



TESIS SS14-2501

**PEMODELAN REGRESI TOBIT KUANTIL BAYESIAN  
PADA PENGELUARAN RUMAH TANGGA  
UNTUK KONSUMSI SUSU**

EVELLIN DEWI LUSIANA

NRP. 1313 201 010

DOSEN PEMBIMBING

Dr. Dra. Ismaini Zain, M.Si

PROGRAM MAGISTER  
JURUSAN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2015



THESIS SS14-2501

# **TOBIT QUANTILE BAYESIAN REGRESSION MODELLING OF HOUSEHOLD EXPENDITURE FOR MILK CONSUMPTION**

**EVELLIN DEWI LUSIANA**

**NRP. 1313 201 010**

**SUPERVISOR**

**Dr. Dra. Ismaini Zain, M.Si**

**MAGISTER PROGRAM  
STATISTICS DEPARTMENT  
FACULTY OF MATHEMATICS AND NATURAL SCIENCES  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2015**

# **PEMODELAN REGRESI TOBIT KUANTIL BAYESIAN PADA PENGELUARAN RUMAH TANGGA UNTUK KONSUMSI SUSU**

**Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar**

**Magister Sains (M.Si)**

**di**

**Institut Teknologi Sepuluh Nopember**

**oleh:**

**EVELLIN DEWI LUSIANA**

**NRP. 1313 201 010**

**Tanggal Ujian: 22 April 2015**

**Periode Wisuda: September 2015**

**Disetujui oleh:**

  
**1. Dr. Dra. Ismaini Zain, M.Si**  
**NIP: 19600525 198803 2 001**

**(Pembimbing)**

  
**2. Dr. Brodjol Sutijo Suprih U., M.Si**  
**NIP: 19660125 199002 1 001**

**(Penguji)**

  
**3. Dr. Vita Ratnasari, S.Si, M.Si**  
**NIP: 19700910 199702 2 001**

**(Penguji)**

**Direktur Program Pascasarjana,**

  
**Prof. Dr. Ir. Adi Soeprijanto, MT**  
**NIP. 19640405 199002 1 001**



# PEMODELAN REGRESI TOBIT KUANTIL BAYESIAN PADA PENGELUARAN RUMAH TANGGA UNTUK KONSUMSI SUSU

Nama Mahasiswa : Evellin Dewi Lusiana  
NRP : 1313201010  
Dosen Pembimbing : Dr. Dra. Ismaini Zain, M.Si

## ABSTRAK

Pemodelan regresi Tobit diterapkan terhadap data tersensor, apabila observasi bernilai nol pada data tersebut juga dipandang sebagai bagian dari proses analisis. Estimasi parameter model regresi Tobit umumnya menggunakan metode *Maximum Likelihood Estimation* (MLE) yang berbasis *conditional mean*, sehingga data yang jauh dari mean, misalnya kuantil 0.05 dan 0.95, seringkali tidak terepresentasi dengan baik oleh estimator yang ada. Hal ini dapat diatasi dengan model regresi kuantil tobit atau regresi kuantil tersensor yang berbasis *conditional quantile*. Estimator model ini dikenal sebagai estimator Powell. Selain itu, diperkenalkan model regresi tobit kuantil dengan pendekatan bayesian (TKB). Kelebihan pendekatan bayesian adalah kemudahan untuk mengestimasi distribusi posterior dengan teknik MCMC (Metropolis-Hastings), serta mampu mengakomodasi adanya informasi prior. Untuk mendapatkan estimator TKB, diasumsikan error model mengikuti distribusi Asimetrik Laplace, sehingga bisa didapatkan fungsi likelihood untuk menghitung posterior. Dari simulasi yang telah dilakukan, diperoleh hasil bahwa estimator TKB dan Powell tidak cukup baik untuk mengestimasi parameter model kuantil bawah seperti kuantil 0.05 dan 0.25. Selain itu, jika prediktor dalam model cukup banyak dan ukuran sampel lebih dari 1000, performa estimator TKB lebih baik daripada estimator Powell. Di samping itu, bila dibandingkan dengan estimator tobit standar, estimator TKB dan Powell memiliki performa yang lebih baik untuk model dengan error berdistribusi bukan normal. Adapun pemodelan regresi TKB terhadap pengeluaran rumah tangga untuk konsumsi susu menunjukkan bahwa besarnya estimator yang dihasilkan bervariasi antar kuantil dan model terbaik untuk masing-masing kuantil juga melibatkan prediktor yang berbeda.

Kata kunci: Tobit, regresi Kuantil, Tobit Kuantil Bayesian (TKB), Metropolis-Hastings, konsumsi susu



# **TOBIT QUANTILE BAYESIAN REGRESSION MODELLING OF HOUSEHOLD EXPENDITURE FOR MILK CONSUMPTION**

Name : Evellin Dewi Lusiana  
NRP : 1313201010  
Supervisor : Dr. Dra. Ismaini Zain, M.Si

## **ABSTRACT**

Tobit regression modelling is applied for censored data, if the null observations are also taken into account. The parameters of tobit regression model are commonly estimated by using Maximum Likelihood Estimation (MLE) or Ordinary Least Square (OLS) method, of which are conditional mean method based, so that the observations that lied afar from mean, such as at 5th and 95th quantile can not well represented. The matter can be solved by using quantile tobit regression or censored quantile regression as the conditional quantile based estimation method, that the estimators are called Powell's estimator. Meanwhile, a tobit quantile model with bayesian approach (TQBR) is also introduced. The advantages of the last model are its ability to obtain the posterior distributions of the parameter of interest even in a complex situations through MCMC method and to accomodate the prior informations. In order to obtain TQBR's estimator, the errors' model are assumed to be Asymmetric Laplace distributed, in case to form the likelihood function for posterior calculation. According to the result of simulation study, both TQBR and Powell estimators are not good to estimate the lower quantile model such as 5th and 25th quantile. Also, their performances are better than standar tobit estimator for model with non-normal error distribution. Specifically, the performance of TQBR's estimators are outperformed the Powell's when the model contains many predictor and the sample size above 1000. On the other hand, the TQBR modelling of household expenditure for milk consumption shows that the estimators are varying across the quantiles and the best model for each quantile involvels different predictors.

**Keywords:** Tobit, Quantile Regression, Tobit Quantile Bayesian Regression (TQBR), Metropolis-Hastings, milk consumption

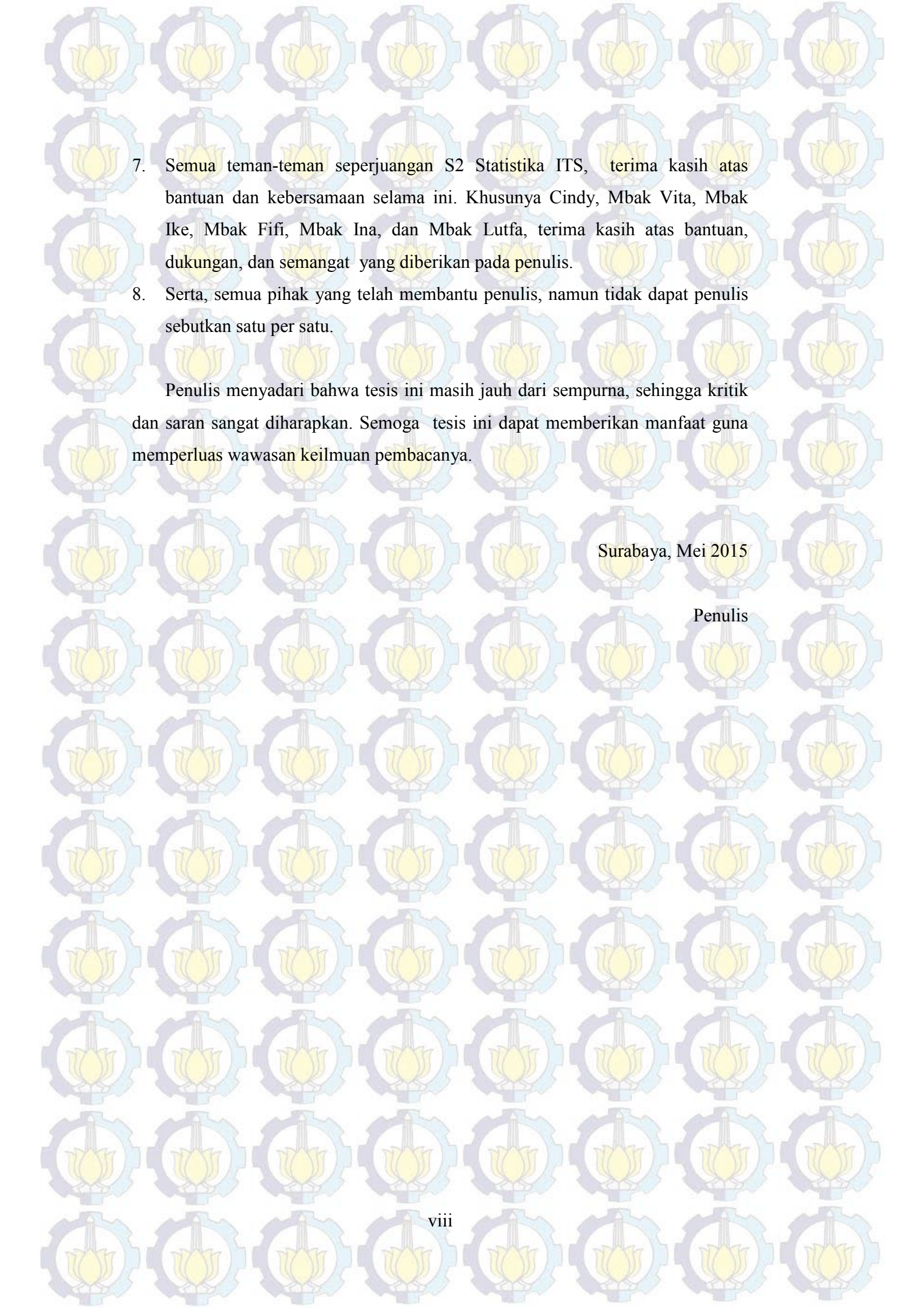


## KATA PENGANTAR

Puji dan syukur penulis hadiratkan kepada Allah SWT, karena atas segala rahmat dan ridho-Nya sehingga tesis yang diberi judul “Pemodelan Regresi Tobit Kuantil Bayesian Pada Pengeluaran Rumah Tangga Untuk Konsumsi Susu” ini bisa terselesaikan. Tesis ini merupakan salah satu syarat untuk menyelesaikan pendidikan di Program Magister S2 Statistika ITS. Ada banyak pihak yang telah membantu dalam penulisan tesis ini, sehingga penulis ingin menyampaikan ucapan terima kasih kepada

1. Ibu Dr. Dra. Ismaini Zain selaku dosen pembimbing, yang telah bersedia meluangkan waktu untuk memberikan bimbingan, saran, dan ilmu yang sangat bermanfaat dalam penyelesaian tesis ini
2. Bapak Dr. Brodjol Sutijo S.U., M.Si dan Ibu Dr. Vita Ratnasari, S.Si, M.Si selaku dosen penguji yang telah memberikan banyak saran dan masukan agar tesis ini menjadi lebih baik
3. Bapak Dr. Muhammad Mashuri, M.T. selaku Ketua Jurusan Statistika ITS dan Bapak Dr. Suhartono, M.Sc selaku Kaprodi Pascasarjana Statistika FMIPA ITS
4. Bapak /Ibu dosen pengajar di Jurusan Statistika ITS, terima kasih atas semua ilmu berharga yang telah diberikan
5. Bapak/Ibu staf dan karyawan di Jurusan Statistika ITS, terima kasih atas segala bantuan selama masa perkuliahan penulis
6. Kedua orang tua yang sangat penulis sayangi dan hormati, Ibu Sumianti dan Bapak Dr. Ir. Sunadji, M.P. Tak ada kalimat yang mampu mengungkapkan besarnya rasa terima kasih penulis kepada mereka atas semua do'a, kasih sayang, kesabaran dan dukungan yang dicurahkan kepada penulis selama ini, juga Dillah, terima kasih telah menjadi adik yang ceria dan selalu membuat penulis tersenyum dan tertawa dengan semua tingkah dan candaan yang dilontarkan. Serta seluruh keluarga besar di Malang, terima kasih atas dukungan yang diberikan



- 
7. Semua teman-teman seperjuangan S2 Statistika ITS, terima kasih atas bantuan dan kebersamaan selama ini. Khususnya Cindy, Mbak Vita, Mbak Ike, Mbak Fifi, Mbak Ina, dan Mbak Lutfa, terima kasih atas bantuan, dukungan, dan semangat yang diberikan pada penulis.
  8. Serta, semua pihak yang telah membantu penulis, namun tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa tesis ini masih jauh dari sempurna, sehingga kritik dan saran sangat diharapkan. Semoga tesis ini dapat memberikan manfaat guna memperluas wawasan keilmuan pembacanya.

Surabaya, Mei 2015

Penulis



## DAFTAR ISI

|   |     |
|---|-----|
| <b>LEMBAR PENGESAHAN</b> .....  | i   |
| <b>ABSTRAK</b> .....  | iii |
| <b>ABSTRACT</b> .....   | v   |
| <b>KATA PENGANTAR</b> .....   | vii |
| <b>DAFTAR ISI</b> .....   | ix  |
| <b>DAFTAR TABEL</b> .....   | xi  |
| <b>DAFTAR GAMBAR</b> .....  | xii |
| <b>DAFTAR LAMPIRAN</b> .....  | xv  |
| <b>BAB 1 PENDAHULUAN</b> .....  | 1   |
| 1.1 Latar Belakang .....  | 1   |
| 1.2 Rumusan Masalah .....   | 4   |
| 1.3 Tujuan Penelitian .....   | 4   |
| 1.4 Manfaat Penelitian .....  | 4   |
| 1.5 Batasan Masalah .....   | 5   |
| <b>BAB 2 TINJAUAN PUSTAKA</b> .....   | 7   |
| 2.1 Regresi Tobit .....   | 7   |
| 2.1.1 Model Regresi Tobit .....   | 7   |
| 2.1.2 Estimasi Parameter Model Regresi Tobit .....                                    | 8   |
| 2.2 Uji Linieritas .....  | 9   |
| 2.3 Uji Beda .....  | 11  |
| 2.4 Model Regresi Kuantil .....   | 12  |
| 2.5 Model Regresi Kuantil Bayesian .....  | 15  |
| 2.5.1 Metode MCMC (Markov Chain Monte Carlo): Algoritma Metropolis-<br>Hastings ..... | 16  |
| 2.5.2 Bayes Faktor .....  | 17  |
| 2.6 <i>Root Mean Square Error (RMSE)</i> .....  | 18  |



|   |           |
|---|-----------|
| 2.7 Faktor-faktor yang Memengaruhi Pengeluaran Rumah Tangga untuk Konsumsi Susu ..... | 18        |
| <b>BAB 3 METODE PENELITIAN .....</b>  | <b>21</b> |
| 3.1 Sumber Data .....   | 21        |
| 3.2 Variabel Penelitian.....  | 21        |
| 3.3 Metode Penelitian .....   | 22        |
| <b>BAB 4 HASIL DAN PEMBAHASAN .....</b>   | <b>29</b> |
| 4.1 Estimasi Parameter dan Pemilihan Model Terbaik Regresi TKB .....                  | 29        |
| 4.1.1 Estimasi Parameter Model Regresi TKB.....                                       | 29        |
| 4.1.2 Pemilihan Model Terbaik Regresi TKB .....                                       | 30        |
| 4.2 Perbandingan Performa Estimator TKB dan Powell .....                              | 31        |
| 4.3 Pemodelan Regresi TKB Bagi Pengeluaran Rumah Tangga untuk Konsumsi Susu .....     | 34        |
| 4.3.1 Deskripsi Data Penelitian .....   | 34        |
| 4.3.2 Hasil Uji Beda Karakteristik Wilayah .....                                      | 36        |
| 4.3.3 Hasil Pengujian Linieritas Variabel Prediktor .....                             | 37        |
| 4.3.4 Hasil Estimasi Parameter Regresi TKB.....                                       | 39        |
| 4.3.5 Pemilihan Model Regresi TKB Terbaik .....                                       | 40        |
| <b>BAB 5 KESIMPULAN DAN SARAN .....</b>   | <b>45</b> |
| 5.1 Kesimpulan .....  | 45        |
| 5.2 Saran .....   | 46        |
| <b>DAFTAR PUSTAKA .....</b>   | <b>49</b> |
| <b>LAMPIRAN - LAMPIRAN .....</b>  | <b>53</b> |
| <b>BIOGRAFI PENULIS.....</b>  | <b>70</b> |



## DAFTAR TABEL

| Judul Tabel   | Halaman |
|---|---------|
| Tabel 2.1 Kriteria Interpretasi Bayes Faktor.....   | 17      |
| Tabel 3.2 Struktur Data Untuk Analisis.....   | 22      |
| Tabel 4.3 Statistik Deskriptif Pengeluaran Rumah Tangga untuk Konsumsi Menurut Wilayah..... | 34      |
| Tabel 4.4 Statistik Deskriptif Variabel Prediktor Penelitian Bersifat Kontinu.....          | 35      |
| Tabel 4.5 Hasil Uji t Karakteristik Wilayah Pedesaan dan Perkotaan.....                     | 37      |
| Tabel 4.6 Hasil Uji <i>chi-square</i> Karakteristik Wilayah Pedesaan dan Perkotaan...       | 37      |
| Tabel 4.7 Hasil Uji RESET.....  | 38      |
| Tabel 4.8 Hasil Uji Korelasi <i>Point Biserial</i> Prediktor <i>Dummy</i> $X_2$ .....       | 38      |
| Tabel 4.9 Hasil Estimasi Parameter Regresi TKB Wilayah Pedesaan.....                        | 39      |
| Tabel 4.10 Hasil Estimasi Parameter Regresi TKB Wilayah Perkotaan .....                     | 40      |
| Tabel 4.11 Bayes Faktor (2 Log B) Model Regresi TKB Reduksi Pedesaan.....                   | 41      |
| Tabel 4.12 Bayes Faktor (2 Log B) Model Regresi TKB Reduksi Perkotaan.....                  | 42      |
| Tabel 4.13 Bayes Faktor Model Regresi TKB Perkotaan Kuantil 0.50.....                       | 43      |



## DAFTAR GAMBAR

| Judul Gambar   | Halaman |
|--|---------|
| Gambar 2.1 <i>Check function</i> $\rho_{\theta}(u) = (\theta - 1_{\{u < 0\}})u$ .....  | 14      |
| Gambar 3.1 Diagram Alir Algoritma Perbandingan Model Regresi Tobit Kuantil Bayesien dengan Model Regresi Kuantil Tersensor ..... | 24      |
| Gambar 3.2 Diagram Alir Algoritma Metropolis-Hastings .....  | 25      |
| Gambar 3.3 Kerangka Penelitian .....   | 27      |



## BAB 1

### PENDAHULUAN

#### 1.1 Latar Belakang

Dalam pemodelan regresi seringkali dijumpai data dengan variabel respon yang terdiri dari observasi bernilai nol dengan proporsi yang cukup besar, sedangkan observasi lainnya memiliki nilai tertentu yang bersifat kontinu. Misal, data mengenai pengeluaran rumah tangga untuk barang-barang mewah dan bahan makanan yang tergolong mahal. Data dengan kondisi semacam ini disebut data tersensor. Jika data tersensor dimodelkan regresi linier, maka observasi-observasi bernilai nol tidak diikutsertakan dalam proses estimasi. Di lain pihak, seandainya digunakan model logistik atau probit, maka variasi dari observasi-observasi yang bernilai tak-nol menjadi tidak diperhitungkan. Oleh karena itu, kedua jenis model ini akan mengakibatkan sebagian informasi dari data tersensor hilang. Keadaan ini dapat diatasi dengan menggunakan model regresi Tobit (Greene, 2001).

Kelebihan dari model regresi Tobit dapat mengakomodasi semua observasi, baik yang bernilai nol maupun tak-nol. Model regresi Tobit sudah cukup luas digunakan seperti Komrattanapanya (2013), Aisyah *et al.* (2012), Zain *et al.* (2011), dan Zain (1997). Komrattanapanya (2013) memanfaatkan regresi Tobit untuk memodelkan pembagian hasil keuntungan saham firma-firma yang terdaftar di bursa saham Thailand. Aisyah *et al.* (2012) menggunakan model regresi Tobit dan *Data Envelopment Analysis* (DEA) untuk mengetahui faktor-faktor yang memengaruhi efisiensi teknis usaha perikanan pantai di Kuala Trengganu, Malaysia. Selain itu, Zain *et al.* (2011) memodelkan data tingkat Pengangguran Terbuka (TPT) perempuan di pulau Jawa menggunakan model tobit yang dengan aspek wilayah. Di sisi lain, Zain (1997) mengkaji model tobit dan aplikasinya pada faktor-faktor yang berpengaruh terhadap pendapatan kerja istri.



Estimasi parameter model regresi Tobit umumnya menggunakan metode *Maximum Likelihood Estimation* (MLE) yang berbasis *conditional mean* yaitu estimator yang diperoleh menitikberatkan pada mean dari distribusi variabel respon. Akibatnya, data yang jauh dari mean, sebagai misal pada kuantil 0.05 dan 0.95, seringkali tidak terepresentasi dengan baik oleh estimator yang ada, sehingga informasi spesifik mengenai data tersebut tidak tereksplorasi dengan baik. Metode lain yang mampu menggambarkan hubungan antara variabel prediktor terhadap variabel respon pada berbagai titik kuantil (*conditional quantile*) adalah analisis regresi kuantil. Analisis ini pertama kali diperkenalkan oleh Koenker dan Basset (1978) dan dapat digunakan pada kondisi data yang heterogen. Estimator parameter model regresi kuantil didapatkan dengan metode pemrograman linier. Model regresi tobit kuantil atau regresi kuantil tersensor diperkenalkan oleh Powell (1986), sehingga dikenal sebagai estimator Powell. Menurut Buchinsky (1994), estimator Powell dapat diperoleh dengan metode *Iterative Linear Programming Algorithm* (ILPA), sedangkan Fitzenberger dan Winker (2007) menggunakan metode BRCENS yang dinyatakan lebih unggul dari ILPA. Di sisi lain, Yu dan Stander (2007) serta Wang dan Zhang (2012) mengusulkan model regresi tobit kuantil dengan pendekatan bayesian. Metode analisis yang menggunakan pendekatan bayesian memiliki keuntungan yaitu kemudahan untuk memperoleh distribusi posterior dari setiap parameter yang ada dengan teknik MCMC. Selain itu, metode ini juga dapat mengakomodasi informasi prior mengenai parameter-parameter yang akan diestimasi, sehingga estimator yang dihasilkan lebih akurat.

Perkembangan model regresi Tobit Kuantil Bayesian (TKB) masih belum banyak dilakukan. Terlihat dari masih sedikit penelitian yang menggunakan model ini, seperti Yue dan Hong (2012) yang memodelkan data pengeluaran rumah tangga untuk kesehatan, serta Yu dan Stander (2007) yang melakukan perbandingan performa estimator parameter model regresi TKB dengan estimator Powell dengan metode ILPA menggunakan teknik simulasi yang mengadopsi Biliyas *et al.* (2000). Hasil perbandingan tersebut menunjukkan bahwa estimator parameter model regresi TKB lebih baik daripada estimator Powell.



Performa estimator parameter model regresi TKB penting untuk didemonstrasikan agar meningkatkan keyakinan para peneliti untuk menggunakan model ini. Menurut Lehmann dan Casella (1999), evaluasi kebaikan suatu estimator dapat dilihat dari perbandingan *performance* (performa) estimator tersebut dengan estimator lain. Salah satu statistik yang umum digunakan untuk membandingkan performa estimator adalah *Root Mean Square Error* (RMSE). Namun, mengingat estimator Powell tidak hanya dapat diperoleh menggunakan metode ILPA, tetapi juga dengan metode BRCENS, sehingga perlu dilakukan perbandingan kembali antara estimator Powell dan regresi TKB.

Model regresi TKB dapat diterapkan pada data pengeluaran rumah tangga untuk konsumsi susu yang dapat diperoleh dari hasil Survei sosial ekonomi nasional (Susenas) dan Survei sosial ekonomi rumah tangga Indonesia (Suseti). Berdasarkan laporan USDA (2013), konsumsi susu per kapita di Indonesia masih di bawah negara-negara lain di kawasan Asia Tenggara dengan hanya sebesar 12.83 liter per tahun, dibandingkan dengan Malaysia (50.9 liter/tahun), Thailand (33.7 liter/tahun) dan Philipina (22.1 liter/tahun). Meskipun rendah, akan tetapi angka ini mengalami peningkatan dari tahun sebelumnya yaitu 11.93 liter/tahun. Peningkatan ini dipicu oleh perkembangan masyarakat kelas menengah, yang semakin sadar akan pentingnya manfaat susu. Keadaan ini menunjukkan bahwa konsumsi susu sangat berkaitan dengan gaya hidup konsumen yang dilatarbelakangi oleh faktor-faktor sosio-ekonomi dan demografis seperti usia, pendapatan, pendidikan, pekerjaan dan budaya (Wham dan Worsley, 2003).

Berdasarkan uraian-uraian yang telah dijabarkan tersebut, penelitian ini akan membahas kajian mengenai estimasi parameter dan metode pemilihan model regresi TKB terbaik. Selain itu, performa estimator parameter model ini juga akan didemonstrasikan melalui perbandingan dengan estimator Powell (metode BRCENS) menggunakan teknik simulasi. Terakhir, model regresi TKB akan diterapkan pada data pengeluaran rumah tangga untuk konsumsi susu.



## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan, dapat dirumuskan beberapa masalah berikut.

1. Bagaimana mengkaji estimator parameter dan pemilihan model regresi TKB terbaik?
2. Bagaimana perbandingan performa estimator parameter model regresi TKB dengan estimator Powell?
3. Bagaimana model regresi TKB terbaik bagi pengeluaran rumah tangga untuk konsumsi susu?

## **1.3 Tujuan Penelitian**

Berikut ini merupakan tujuan yang ingin dicapai dalam penelitian ini.

1. Mengkaji estimasi parameter dan pemilihan model regresi TKB terbaik
2. Membandingkan performa estimator parameter model regresi TKB dengan estimator Powell menggunakan teknik simulasi.
3. Memodelkan regresi TKB terbaik bagi pengeluaran rumah tangga untuk konsumsi susu.

## **1.4 Manfaat Penelitian**

Manfaat dari hasil penelitian ini nantinya sebagai berikut.

1. Mengembangkan wawasan dan pengetahuan mengenai analisis regresi kuantil pada umumnya dan regresi TKB pada khususnya serta pemilihan model terbaiknya.
2. Memperoleh informasi mengenai seberapa baik performa dari model regresi TKB bila digunakan dalam berbagai kondisi data yang tersensor dan dibandingkan dengan estimator Powell.
3. Memberikan informasi yang lebih lengkap tentang model pengeluaran rumah tangga untuk konsumsi susu, sehingga diharapkan bisa membantu upaya pengambil kebijakan untuk meningkatkan konsumsi susu di suatu wilayah.





### 1.5 Batasan Masalah

Beberapa batasan masalah dalam penelitian ini adalah sebagai berikut.

1. Ukuran untuk mengevaluasi performa estimator adalah RMSE.
2. Estimator Powell yang akan dibandingkan dengan estimator parameter model regresi tobit kuantil bayesian diperoleh menggunakan metode BRCENS.
3. Kuantil yang akan dimodelkan adalah 0.05, 0.25, 0.50, 0.75 dan 0.95.



## BAB 2

### TINJAUAN PUSTAKA

#### 2.1 Regresi Tobit

##### 2.1.1 Model Regresi Tobit

Pada awal kemunculan model tobit diperuntukan untuk menganalisis pengeluaran rumah tangga yang seringkali bernilai nol untuk barang-barang tertentu yang tergolong mewah, atau bisa juga untuk permasalahan di mana terjadi perubahan kondisi rumah tangga bila melewati batasan (*limit*) tertentu yang bukan nol. Apabila variabel semacam ini dijadikan sebagai variabel respon dalam suatu analisis regresi, maka keberadaan nilai *limit* tersebut harus mendapat perhatian ketika dilakukan estimasi parameter regresi antar variabel ini dengan variabel-variabel prediktor serta pengujian hipotesis atas hubungan kedua variabel tersebut. Apabila besarnya nilai observasi variabel respon yang tidak masuk dalam *limit* diabaikan, maka model regresi probit layak untuk digunakan. Meskipun demikian, akan sangat merugikan untuk membuang informasi mengenai nilai-nilai variabel respon yang tidak masuk dalam *limit*, ketika informasi tersebut tersedia. Oleh karena itu, diusulkanlah penggunaan model regresi tobit untuk mengatasi hal tersebut (Tobin, 1958).

Menurut McBee (2010), misalkan terdapat sebanyak  $n$  data observasi yang terdiri atas satu variabel respon ( $Y$ ) dan  $p$  variabel prediktor ( $X_1, X_2, \dots, X_p$ ), maka variabel respon  $Y$  dikatakan tersensor pada batas bawah apabila untuk setiap  $i=1,2,\dots,n$  berlaku persamaan (2.1).

$$y_i = \begin{cases} \tau & y_i^* \leq \tau \\ y_i^* & y_i^* > \tau \end{cases} \quad (2.1)$$

Leiker (2012) menyatakan bahwa persamaan (2.1) dapat juga ditulis menjadi persamaan (2.2).

$$y_i = \max(y_i^*, \tau) \quad (2.2)$$



Nilai observasi  $y_i^*$  dapat dinyatakan sebagai persamaan (2.3).

$$y_i^* = f(\mathbf{x}_i) + \varepsilon_i \quad (2.3)$$

di mana:  $f(\mathbf{x}_i) = \mathbf{x}_i' \boldsymbol{\beta}$

Dengan demikian, diperoleh model regresi tobit dalam persamaan (2.4) atau (2.5).

$$y_i = \begin{cases} \tau & y_i^* \leq \tau \\ \mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i & y_i^* > \tau \end{cases} \quad (2.4)$$

$$y_i = \max(\mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tau) \quad (2.5)$$

di mana:

$y_i$  = nilai variabel respon tersensor ke- $i$

$y_i^*$  = nilai variabel respon laten ke- $i$

$\tau$  = nilai titik sensor yang diketahui

$\mathbf{x}_i = (1, x_{1i}, x_{2i}, \dots, x_{pi})$

$\boldsymbol{\beta}$  = parameter model regresi tobit

$\varepsilon_i$  = error model yang berdistribusi  $N(0, \sigma^2)$

$i = 1, 2, \dots, n$

Apabila model regresi tobit dengan variabel respon  $Y$  tersensor pada batas atas, maka tanda pertidaksamaan pada persamaan (2.4) dan (2.5) diubah sebaliknya.

### 2.1.2 Estimasi Parameter Model Regresi Tobit

Parameter model regresi tobit dapat diestimasi dengan metode *Maximum Likelihood Estimation* (MLE). Fungsi likelihood untuk model regresi tobit dibedakan menjadi dua yaitu untuk  $y_i^* > \tau$  dan  $y_i^* \leq \tau$ . Jika  $\tau=0$ , fungsi likelihood model regresi tobit dapat dituliskan sebagai persamaan (2.6) yang merupakan gabungan dua fungsi (Wooldridge, 2002).



$$L(\boldsymbol{\beta}) = \prod_{i=1}^n \left[ 1(y_i > 0) \cdot \frac{1}{\sigma} \phi \left( \frac{y_i - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right] \left[ 1(y_i = 0) \left\{ 1 - \Phi \left( \frac{\mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right\} \right] \quad (2.6)$$

$$l(\boldsymbol{\beta}) = \log L(\boldsymbol{\beta}) = \sum_{i=1}^n 1(y_i > 0) \cdot \log \left\{ \frac{1}{\sigma} \phi \left( \frac{y_i - \mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right\} + 1(y_i = 0) \cdot \log \left\{ 1 - \Phi \left( \frac{\mathbf{x}_i' \boldsymbol{\beta}}{\sigma} \right) \right\} \quad (2.7)$$

Estimator parameter model regresi tobit diperoleh dari turunan pertama persamaan (2.7) terhadap  $\boldsymbol{\beta}$  yang disamadengkan nol dengan mengasumsikan asumsikan  $\sigma^2$  diketahui. Namun karena penyelesaian persamaan ini bersifat nonlinier, maka sulit dilakukan perhitungan secara analitis. Oleh sebab itu, digunakan metode iteratif Newton-Raphson (Greene, 2001).

## 2.2 Uji Linieritas

Berdasarkan persamaan (2.3), variabel respon laten  $y_i^*$  dapat ditulis dalam bentuk model yang menyerupai model regresi linier berganda. Menurut Osborne dan Waters (2002), parameter model regresi linier berganda hanya bisa dengan tepat terestimasi apabila hubungan antara prediktor dan respon bersifat linier. Jika hubungan keduanya nonlinier, maka akan menimbulkan *under-estimate*. Salah satu cara untuk menguji linieritas hubungan prediktor dan respon adalah dengan menguji komponen nonlinier yang dimasukkan ke dalam model, seperti uji RESET.

Misal, didefinisikan model regresi primer sesuai (2.3), tahap pertama dalam uji RESET adalah mendapatkan nilai duga  $\hat{y}_i$ . Selanjutnya, nilai duga tersebut dimasukkan ke model regresi *auxiliary* (2.8) yang mengandung H komponen nonlinier (Shukur dan Mantalos, 2004).

$$y_i = \mathbf{x}_i' \boldsymbol{\beta} + \hat{y}_i^2 \gamma_1 + \hat{y}_i^3 \gamma_2 + \dots + \hat{y}_i^{H+1} \gamma_H + \delta_i \quad (2.8)$$

Hipotesis yang diuji adalah sebagai berikut.

$$H_0 : \gamma_1 = \gamma_2 = \dots = \gamma_H = 0$$

$$H_1 : \text{minimal terdapat satu } \gamma_h \neq 0 \text{ (h=1,2,...,H)}$$

Statistik uji RESET ditunjukkan oleh persamaan (2.9) berikut.



$$RESET = \frac{(\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \boldsymbol{\delta}'\boldsymbol{\delta}) / p}{(\boldsymbol{\delta}'\boldsymbol{\delta}) / (n - H)} \sim F_{(p, n-H-1)} \quad (2.9)$$

di mana:

$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}$  = jumlah kuadrat residual model regresi primer

$\boldsymbol{\delta}'\boldsymbol{\delta}$  = jumlah kuadrat residual model regresi *auxiliary*

$H_0$  ditolak jika  $P(F_{(p, n-H-1)} > RESET) < \alpha$ .

Di sisi lain, apabila terdapat variabel prediktor bersifat kategorik (*dummy*), maka uji yang digunakan adalah uji korelasi *Point Biserial*. Uji ini bertujuan untuk mengetahui adanya hubungan linier antar dua variabel di mana salah satu variabel bersifat biner/dikotomis dan variabel lainnya kontinu. Hipotesis yang diuji sebagai berikut.

$H_0$  : hubungan antar variabel tidak linier

$H_1$  : hubungan antar variabel linier

Menurut Glass dan Hopkins (1995), perhitungan korelasi *Point biserial* ditunjukkan oleh persamaan (2.10), sedangkan statistik uji untuk korelasi dapat dituliskan sebagai persamaan (2.11).

$$r_{pb} = \frac{\bar{y}_1 - \bar{y}_0}{s_y} \sqrt{\frac{c_1 c_0}{c(c-1)}} \quad (2.10)$$

$$t = \frac{r_{pb} \sqrt{c-2}}{\sqrt{1-r_{pb}^2}} \sim t_{(c-2)} \quad (2.11)$$

di mana:

$\bar{y}_1$  = rata-rata respon kategori 1

$\bar{y}_0$  = rata-rata respon kategori 0

$c_1$  = banyaknya observasi kategori 1

$c_0$  = banyaknya observasi kategori 0

$c$  =  $c_1 + c_0$

$s_y$  = standar deviasi variabel respon y

$H_0$  ditolak jika  $P(t_{(c-2)} > |t|) < \alpha$ .



### 2.3 Uji Beda

Menurut Robinson *et al.* (2013), uji beda dapat dilakukan untuk mengetahui pengaruh suatu variabel yang berpotensi menjadi moderator dalam suatu analisis regresi. Apabila variabel yang kemungkinan adalah moderator bersifat biner, dan pengaruhnya diuji terhadap variabel lain yang bersifat kontinu, maka uji beda yang digunakan adalah uji t. Hipotesis yang diuji dengan uji t sebagai berikut.

$$H_0 : \mu_1 - \mu_0 = 0$$

$$H_1 : \mu_1 - \mu_0 \neq 0$$

Bentuk umum statistik uji t ditunjukkan oleh persamaan (2.12).

$$t = \frac{\bar{x}_1 - \bar{x}_0}{s_{\bar{x}_1 - \bar{x}_0}} \sim t_{(db)} \quad (2.12)$$

Jika ragam kelompok sama, maka  $s_{\bar{x}_1 - \bar{x}_0} = \sqrt{\frac{(n_0 - 1)s_{x_0}^2 + (n_1 - 1)s_{x_1}^2}{n_0 + n_1 - 2}}$  dan

$db = n_0 + n_1 - 2$ . Di sisi lain, jika ragam kelompok berbeda, digunakan

$$s_{\bar{x}_1 - \bar{x}_0} = \sqrt{\frac{s_{x_0}^2}{n_0} + \frac{s_{x_1}^2}{n_1}} \text{ dan db efektif.}$$

di mana:

$\bar{x}_1$  = rata-rata sampel kelompok kategori 1

$\bar{x}_0$  = rata-rata sampel kelompok kategori 0

$$s_{\bar{x}_1 - \bar{x}_0} = \begin{cases} \sqrt{s_p^2 \left( \frac{1}{n_0} + \frac{1}{n_1} \right)} & , \text{ jika ragam sama} \\ \sqrt{\frac{s_{x_0}^2}{n_0} + \frac{s_{x_1}^2}{n_1}} & , \text{ jika ragam berbeda} \end{cases}$$

$$s_p^2 = \sqrt{\frac{(n_0 - 1)s_{x_0}^2 + (n_1 - 1)s_{x_1}^2}{n_0 + n_1 - 2}}$$



$$db = \begin{cases} n_0 + n_1 - 2 & , \text{jika ragam sama} \\ \frac{\left( \frac{s_{x_0}^2}{n_0} + \frac{s_{x_1}^2}{n_1} \right)^2}{\left( \frac{s_{x_0}^2}{n_0} \right)^2 / (n_0 - 1) + \left( \frac{s_{x_1}^2}{n_1} \right)^2 / (n_1 - 1)} & , \text{jika ragam berbeda} \end{cases}$$

$n_0$  = banyaknya observasi kelompok 0

$n_1$  = banyaknya observasi kelompok 1

$H_0$  ditolak jika  $P(t_{(db)} > |t|) < \alpha/2$ .

Apabila kedua variabel sama-sama bersifat biner, maka dipakai uji independensi *chi-square*. Uji ini didasarkan pada tabel kontingensi 2x2 dan sering pula disebut uji kebebasan *chi-square*. Adapun hipotesis untuk uji *chi-square* sebagai berikut.

$H_0$  : hubungan antar kedua variabel saling bebas

$H_1$  : hubungan antar kedua variabel tidak saling bebas

Persamaan (2.13) berikut adalah statistik uji *chi-square* (Agresti, 2002).

$$\chi^2 = \sum_{m=1}^g \frac{(O_m - E_m)^2}{E_m} \sim \chi_{(g-1)}^2 \quad (2.13)$$

di mana:

$O_m$  = frekuensi observasi dalam kategori ke-m

$E_m$  = frekuensi yang diharapkan dalam kategori ke-m

$g$  = banyaknya kategori

$H_0$  ditolak jika  $P(\chi_{(g-1)}^2 > \chi^2) < \alpha$ .

## 2.4 Model Regresi Kuantil

Regresi kuantil adalah salah satu metode analisis regresi yang dapat menggambarkan hubungan satu atau beberapa variabel prediktor terhadap satu variabel respon pada berbagai titik kuantil (*conditional quantile*) dari distribusi variabel respon tersebut, sehingga metode ini dapat digunakan pada kondisi data yang heterogen. Hal ini berbeda dengan analisis regresi linier yang hanya dapat menggambarkan hubungan sebab-akibat pada mean (*conditional mean*) variabel respon (Koenker dan Hallock, 2001).



Persamaan (2.14) berikut menunjukkan bentuk umum model regresi kuantil linier (Buhai, 2005).

$$y_i = \mathbf{x}_i' \boldsymbol{\beta}(\theta) + \varepsilon(\theta)_i \quad 0 < \theta < 1 \quad (2.14)$$

di mana:

$y_i$  = nilai variabel respon ke- $i$

$$\mathbf{x}_i' = (1, x_{1i}, x_{2i}, \dots, x_{pi})$$

$\boldsymbol{\beta}(\theta)$  = parameter model regresi pada kuantil ke- $\theta$

$\varepsilon(\theta)_i$  = error model regresi kuantil ke- $\theta$

$i = 1, 2, \dots, n$

Berdasarkan Chen (2005), estimasi parameter model regresi kuantil diawali dengan menyatakan fungsi peluang kumulatif dari variabel random  $Y$  seperti persamaan (2.15), sehingga kuantil ke  $\theta$  dari variabel ini dapat dituliskan sebagaimana persamaan (2.16).

$$F(y) = P(Y \leq y) \quad (2.15)$$

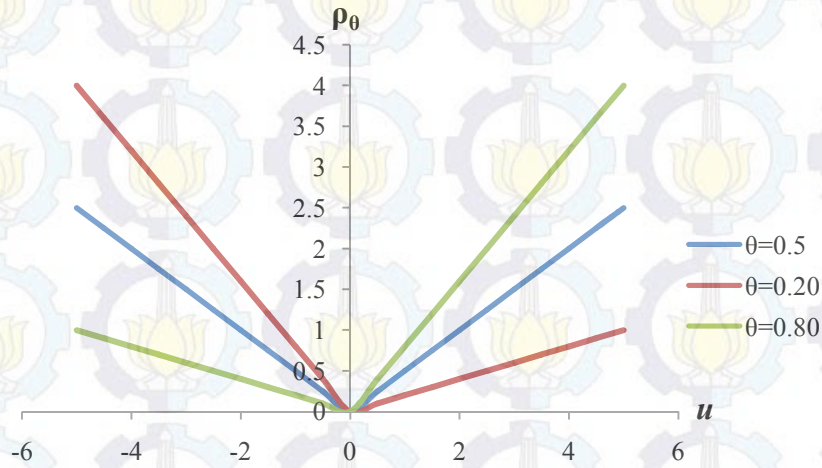
$$Q_Y(\theta) = \inf\{y : F(y) \geq \theta\} \quad (2.16)$$

Menurut Koenker dan Basset (1978), jika terdapat sebanyak  $n$  observasi  $\{y_i : i = 1, \dots, n\}$  sebagai sampel random dari variabel  $Y$  dengan fungsi distribusi  $F$ , maka kuantil ke- $\theta$  dapat didefinisikan sebagai penyelesaian atau solusi dari masalah minimasi persamaan (2.17).

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left[ \sum_{i \in \{i: y_i \geq \mathbf{x}_i' \boldsymbol{\beta}\}} \rho_{\theta}(y_i - f(\mathbf{x}_i)) \right] \quad (2.17)$$

$\rho_{\theta}(u) = (\theta - 1_{\{u < 0\}})u$  merupakan *check function* yang dapat diilustrasikan dengan Gambar 2.1.





Gambar 2.1 Check function  $\rho_{\theta}(u) = (\theta - 1_{\{u < 0\}})u$

Koenker dan Machado (1999) menyatakan bahwa persamaan (2.17) tidak memiliki bentuk turunan yang tetap, sehingga metode iterasi numerik biasa tidak dapat digunakan untuk menyelesaikan persamaan tersebut. Oleh karena itu, untuk mendapatkan  $\hat{\beta}(\theta)$  digunakan metode pemrograman linear yaitu metode simpleks. Formulasi pemrograman linier model regresi kuantil berdasarkan Yao dan Lee (2010) dapat dilihat pada Lampiran 1. Estimasi Model Regresi Kuantil Tersensor

Menurut Powell (1986), model regresi kuantil tersensor untuk kuantil  $\theta$  merupakan hasil substitusi persamaan (2.5) ke (2.17). Oleh karena itu, estimator model ini adalah penyelesaian masalah minimasi (2.18).

$$\min_{\beta \in \mathbb{R}^p} \left[ \sum_{i=1}^n \rho_{\theta}(y_i - \text{maks}(\mathbf{x}_i^T \beta(\theta), \tau)) \right] \quad (2.18)$$

Solusi minimasi (2.18) diperoleh dengan bantuan metode pemrograman linier. Buchinsky (1994) menawarkan penggunaan metode *Iterative Linear Programming Algorithm* (ILPA) untuk menyelesaikan minimasi ini. Namun, menurut Fitzenberger dan Winker (2007) kekurangan metode ILPA adalah tidak ada kepastian konvergensi tercapai dan sekalipun tercapai, hal ini juga tidak



menjamin solusi yang dihasilkan merupakan *local minima* dari (2.18). Oleh karena itu, Fitzenberger dan Winker (2007) menyarankan metode BRCENS yang merupakan pengembangan algoritma Barrodale-Robert untuk menyelesaikan (2.18). Algoritma metode BRCENS dapat ditemukan dalam Fitzenberger (1997).

## 2.5 Model Regresi Kuantil Bayesian

Berdasarkan persamaan (2.17) yang memuat *check function* atau *loss function*, Yu dan Moyeed (2001) menemukan bahwa masalah minimasi pada (2.17) ekuivalen dengan memaksimalkan likelihood dari fungsi asimetris Laplace. Suatu variabel random  $U$  dikatakan berdistribusi asimetris Laplace dengan fungsi peluang (2.19).

$$f_{\theta}(u) = \theta(1 - \theta) \exp(-\rho_{\theta}(u)) \quad , 0 < \theta < 1 \quad (2.19)$$

Apabila diasumsikan residual model regresi kuantil berdistribusi asimetris Laplace, maka persamaan (2.19) menjadi persamaan (2.20).

$$f_{\theta}(\varepsilon_i) = \theta(1 - \theta) \exp(-\rho_{\theta}(y_i - \mathbf{x}_i' \boldsymbol{\beta}(\theta))) \quad , 0 < \theta < 1 \quad (2.20)$$

Menurut Yu dan Moyeed (2001), prinsip dasar pemodelan dengan pendekatan Bayesian adalah mendapatkan distribusi posterior dari suatu parameter bila diketahui distribusi prior dan fungsi likelihood yang sesuai dengan kaidah Bayes. Apabila pendekatan ini digunakan untuk model regresi kuantil, maka distribusi posterior  $\boldsymbol{\beta}(\theta)$  yakni  $\pi(\boldsymbol{\beta}(\theta)|\mathbf{y})$  dapat dirumuskan pada (2.21).

$$\pi(\boldsymbol{\beta}(\theta)|\mathbf{y}) \propto L(\mathbf{y}|\boldsymbol{\beta}(\theta))\pi(\boldsymbol{\beta}(\theta)) \quad (2.21)$$

$\pi(\boldsymbol{\beta}(\theta))$  merupakan distribusi prior dari  $\boldsymbol{\beta}(\theta)$  dan  $L(\mathbf{y}|\boldsymbol{\beta}(\theta))$  adalah fungsi likelihood data dengan asumsi residual berdistribusi asimetris Laplace.

$$L(\mathbf{y}|\boldsymbol{\beta}(\theta)) = \theta^n (1 - \theta)^n \exp \left[ - \sum_{i=1}^n \rho_{\theta}(y_i - \mathbf{x}_i' \boldsymbol{\beta}(\theta)) \right]$$



Distribusi prior yang digunakan oleh Yu dan Moyeed (2001) adalah prior *improper* uniform. Prior ini dipilih karena model regresi kuantil bayesian tidak memiliki prior konjugasi. Selain itu melalui pembuktian secara matematis, diketahui bahwa meskipun prior ini *improper* namun dapat menghasilkan posterior yang bersifat *proper*.

### 2.5.1 Metode MCMC (Markov Chain Monte Carlo): Algoritma Metropolis-Hastings

MCMC adalah metode umum yang digunakan untuk mendapatkan suatu nilai (sampel)  $\beta$  dari distribusi peluang yang diketahui ( $\pi(\beta)$ ), lalu nilai yang didapatkan tersebut dikoreksi sehingga bisa lebih mendekati distribusi posterior yang dikehendaki ( $\pi(\beta|y)$ ). Sampel-sampel  $\beta$  didapatkan secara berurutan (*sequentially*), atau dengan kata lain distribusi peluang yang digunakan untuk menghasilkan sampel saat ini tergantung dari sampel yang dihasilkan sebelumnya, sehingga membentuk suatu rantai Markov (Gelman *et al.*, 2004).

Algoritma berdasarkan prinsip MCMC dapat dibedakan menjadi dua macam yaitu Gibbs sampler dan Metropolis-Hastings. Dalam penelitian ini digunakan algoritma Metropolis-Hastings dengan bentuk umum sebagai berikut (Wang dan Zhang, 2012).

1. Menentukan nilai awal  $\beta^{(1)}$
2. Untuk  $i=2,3,\dots,N$  ( $N$ =banyaknya *loop chain*)
  - a. Tentukan  $\beta = \beta^{(i-1)}$
  - b. Membangkitkan kandidat nilai parameter baru  $\beta'$  dari distribusi proposal (misal Gaussian)  $\pi(\beta'|\beta)$
  - c. Hitung peluang *acceptance*  $\alpha = \min\left(1, \frac{L(y|\beta')\pi(\beta')}{L(y|\beta)\pi(\beta)}\right)$
  - d. Perbarui  $\beta^{(i)} = \beta'$  dengan peluang *acceptance* sebesar  $\alpha$



### 2.5.2 Bayes Faktor

Bayes faktor pertama kali diperkenalkan oleh Harold Jeffreys pada tahun 1960 sebagai alternatif dari pengujian hipotesis frekuentis. Statistik ini sering digunakan untuk membandingkan beberapa model guna mendapatkan model terbaik (Lavine dan Schervish, 1999).

Persamaan (2.22) menunjukkan perbandingan dua model menggunakan Bayes faktor menurut Kass dan Raftery (1995).

$$B = \frac{m(\mathbf{y}|M_1)}{m(\mathbf{y}|M_2)} = \frac{\int L(\mathbf{y}|M_1, \boldsymbol{\beta}_1) \pi(\boldsymbol{\beta}_1|M_1) d\boldsymbol{\beta}_1}{\int L(\mathbf{y}|M_2, \boldsymbol{\beta}_2) \pi(\boldsymbol{\beta}_2|M_2) d\boldsymbol{\beta}_2} \quad (2.22)$$

di mana:

$M_l$  = model ke- $l$  ( $l=1,2$ )

$\mathbf{y}$  =  $(y_1, y_2, \dots, y_n)$

$\boldsymbol{\beta}_l$  = vektor parameter model ke- $l$

$m(\mathbf{y}|M_l)$  = fungsi likelihood marginal model ke- $l$

$L(\mathbf{y}|M_l, \boldsymbol{\beta}_l)$  = fungsi likelihood model ke- $l$

$\pi(\boldsymbol{\beta}_l|M_l)$  = fungsi distribusi prior model ke- $l$

Jika digunakan hipotesis  $M_1 > M_2$ , maka interpretasi Bayes faktor dapat dijelaskan seperti pada Tabel 2.1 (Kass dan Raftery, 1995).

Tabel 2.1 Kriteria Interpretasi Bayes Faktor

| <b>2 log B</b> | <b>B</b> | <b>Kekuatan Pembuktian</b> | <b>Model</b> |
|----------------|----------|----------------------------|--------------|
| < 0            | < 1      | Negatif                    | $M_1$        |
| 0 - 2          | 1 - 3    | Tidak ada                  |              |
| 2 - 5          | 3 - 12   | Positif                    |              |
| 5 - 10         | 12 - 150 | Kuat                       |              |
| >10            | >150     | Sangat kuat                |              |



Kass dan Raftery (1995) menyatakan bahwa perhitungan Bayes faktor cukup sulit untuk dilakukan. Namun, Chib dan Jeliakoz (2001) memberikan gagasan mereka untuk menghitung estimasi logaritma marginal likelihood,  $\log m(\mathbf{y}|M_l)$ , untuk sembarang nilai  $\beta_l^*$  yang memiliki densitas tinggi.

$$\log m(\mathbf{y}|M_l) = \log L(\mathbf{y}|M_l, \beta_l^*) + \log \pi(\beta_l^*|M_l) - \log \pi(\beta_l^*|\mathbf{y}, M_l) \quad (2.23)$$

Persamaan (2.23) dapat dihitung dengan menggunakan output dari algoritma Metropolis-Hastings.

## 2.6 *Root Mean Square Error (RMSE)*

RMSE (*Root Mean Square Error*) merupakan salah satu statistik yang sering digunakan untuk mengevaluasi kebaikan performa model atau estimator. Statistik ini mengukur selisih antara nilai yang diprediksi oleh suatu model/estimator dengan nilai sebenarnya, yang disebut juga sebagai error atau residual. Persamaan (2.24) berikut menunjukkan formulasi RMSE jika diasumsikan terdapat sebanyak  $n$  error model (Chai dan Draxler, 2014).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \quad (2.24)$$

## 2.7 **Faktor-faktor yang Memengaruhi Pengeluaran Rumah Tangga untuk Konsumsi Susu**

Menurut Sukirno (2004), konsumsi adalah pengeluaran total untuk memperoleh barang dan jasa dalam suatu perekonomian dalam jangka waktu tertentu (biasanya 1 tahun) pengeluaran. pengeluaran konsumsi terdiri atas konsumsi pemerintah dan konsumsi rumah tangga. Dari kedua jenis konsumsi ini, pengeluaran konsumsi rumah tangga memiliki porsi terbesar dalam total pengeluaran agregat. Selain itu, konsumsi rumah tangga bersifat endogenus, artinya besarnya konsumsi rumah tangga berkaitan erat dengan faktor – faktor lain yang dianggap memengaruhinya.

Faktor-faktor yang memengaruhi tingkat konsumsi rumah tangga antara lain faktor ekonomi (pendapatan, kekayaan, tingkat bunga dan perkiraan tentang masa depan), faktor demografi (jumlah dan komposisi penduduk) dan faktor non-ekonomi (sosial budaya). Salah satu model teori konsumsi yang cukup populer



adalah model Keynesian. Model ini menjelaskan bahwa konsumsi saat ini (*current consumption*) sangat dipengaruhi oleh pendapatan disposabel saat ini (*current disposable income*). Jika pendapatan disposabel meningkat, maka konsumsi juga akan meningkat (Rahardja dan Manurung, 2008).

Pengeluaran konsumsi rumah tangga untuk susu sangat dipengaruhi oleh faktor-faktor tertentu seperti kondisi wilayah, sosial-budaya, dan sebagainya. Beberapa penelitian menyangkut pengeluaran konsumsi pangan seperti yang dilakukan Sekhampu (2012) serta Gheblawi dan Sherif (2007) menghasilkan kesimpulan bahwa faktor-faktor sosio-ekonomi sangat menentukan keputusan pembelian oleh konsumen. Selain itu, perilaku konsumsi rumah tangga sangat tergantung dari profil demografinya seperti usia kepala rumah tangga, jumlah anggota rumah tangga dan jumlah anggota rumah tangga yang bekerja.

Penelitian Phuong *et al.* (2013) menemukan faktor-faktor sosio-ekonomi dan demografis seperti usia, pendapatan, pendidikan, gender, etnis dan jumlah anak usia kurang dari 12 tahun sangat mempengaruhi pengeluaran rumah tangga untuk produk-produk susu di Vietnam. Tingginya pendapatan, tingkat pendidikan dan kehadiran anak-anak usia  $\leq 12$  tahun akan berpengaruh positif terhadap pengeluaran konsumsi rumah tangga. Hal ini juga sejalan dengan yang dihasilkan oleh Trung *et al.* (2014), Uzunoğlu dan Akcay (2012), dan Alwis *et al.* (2009).

Babolian dan Karim (2010) melakukan penelitian untuk mengetahui pengaruh faktor lingkungan terhadap minat anak-anak untuk mengkonsumsi susu di Selangor-Malaysia menurut wilayah perkotaan dan pedesaan. Hasil penelitian ini menunjukkan bahwa anak-anak yang tinggal di wilayah perkotaan mengkonsumsi susu lebih banyak dibanding mereka yang tinggal di wilayah pedesaan.

Berdasarkan uraian mengenai penemuan beberapa peneliti tersebut, maka penelitian ini akan menggunakan variabel-variabel prediktor yang terdiri atas faktor-faktor sosial-ekonomi yaitu pendapatan rumah tangga, rata-rata pengeluaran rumah tangga, dan persentase pengeluaran rumah tangga untuk pangan, serta faktor-faktor demografi yakni tingkat pendidikan kepala rumah tangga, jumlah anggota rumah tangga, persentase anggota rumah tangga bekerja, dan persentase anggota rumah tangga berusia  $\leq 12$  tahun.



## BAB 3

### METODE PENELITIAN

#### 3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data pengeluaran rumah tangga untuk konsumsi susu yang diperoleh dari hasil Survei Sosial Ekonomi Rumah Tangga Indonesia (SUSETI) tahun 2011. Unit sampel dalam survei ini adalah rumah tangga. Jumlah sampel yang digunakan sebanyak 5,432 unit.

#### 3.2 Variabel Penelitian

Adapun variabel-variabel dalam penelitian ini dapat dibedakan menjadi dua macam yaitu variabel respon dan prediktor sebagai berikut.

1. Variabel respon, yakni pengeluaran suatu rumah tangga untuk konsumsi susu selama satu bulan dalam ribuan rupiah.

$y_i = 0$  jika rumah tangga tidak mengeluarkan biaya untuk konsumsi susu

$y_i = y_i^*$  jika rumah tangga mengeluarkan biaya untuk konsumsi susu

2. Variabel prediktor, dalam penelitian ini digunakan tujuh variabel prediktor yakni

- a.  $X_1$  = Pendapatan rumah tangga, adalah data pendapatan rumah tangga selama satu bulan dalam ribuan rupiah.

- b.  $X_2$  = Tingkat pendidikan kepala rumah tangga, merupakan ijazah tertinggi yang dimiliki oleh kepala rumah tangga, meliputi SD atau tidak tamat SD, SLTP dan SMU, serta Perguruan Tinggi. Oleh karena itu, variabel ini diubah menjadi variabel *dummy* dengan struktur sebagai berikut.

$(X_{21}, X_{22}) = (0,0)$  untuk SD atau tidak tamat SD

$= (1,0)$  untuk SLTP dan SMU

$= (0,1)$  untuk Perguruan Tinggi



- c.  $X_3$  = Persentase pengeluaran makanan/pangan, yaitu jumlah pengeluaran rumah tangga untuk konsumsi makanan dibagi total pengeluaran rumah tangga dan dikali seratus.
- d.  $X_4$  = Jumlah anggota rumah tangga, adalah semua orang yang biasanya tinggal di dalam suatu rumah tangga.
- e.  $X_5$  = Persentase anggota rumah tangga yang bekerja, merupakan banyaknya anggota rumah tangga yang bekerja dibagi dengan jumlah anggota rumah tangga, lalu dikali seratus.
- f.  $X_6$  = Persentase anggota rumah tangga usia  $\leq 12$  tahun, yaitu banyaknya anggota rumah tangga yang berusia 12 tahun atau kurang dibagi jumlah anggota rumah tangga dan dikali seratus.
- g.  $X_7$  = Rata-rata pengeluaran per kapita, merupakan total pengeluaran rumah tangga selama satu bulan dibagi jumlah anggota rumah tangga dalam ribuan rupiah.
- h.  $Z$  = Variabel moderator daerah tempat tinggal, yaitu letak tempat tinggal suatu rumah tangga di perkotaan atau pedesaan.

Adapun struktur data yang digunakan ditampilkan pada Tabel 3.1.

Tabel 3.2 Struktur Data Untuk Analisis

| Sampel ke- | Y        | $X_1$     | $X_{21}$   | $X_{22}$   | $X_3$     | $X_4$     | $X_5$     | $X_6$     | $X_7$     |
|------------|----------|-----------|------------|------------|-----------|-----------|-----------|-----------|-----------|
| 1          | $Y_1$    | $X_{11}$  | $X_{211}$  | $X_{221}$  | $X_{31}$  | $X_{41}$  | $X_{51}$  | $X_{61}$  | $X_{71}$  |
| 2          | $Y_2$    | $X_{12}$  | $X_{212}$  | $X_{222}$  | $X_{32}$  | $X_{42}$  | $X_{52}$  | $X_{62}$  | $X_{72}$  |
| 3          | $Y_3$    | $X_{13}$  | $X_{213}$  | $X_{223}$  | $X_{33}$  | $X_{43}$  | $X_{53}$  | $X_{63}$  | $X_{73}$  |
| ...        | ...      | ...       | ...        | ...        | ...       | ...       | ...       | ...       | ...       |
| $n_t$      | $Y_{nt}$ | $X_{1nt}$ | $X_{21nt}$ | $X_{22nt}$ | $X_{3nt}$ | $X_{4nt}$ | $X_{5nt}$ | $X_{6nt}$ | $X_{7nt}$ |

$t=1 \rightarrow$  Pedesaan;  $t=2 \rightarrow$  Perkotaan

### 3.3 Metode Penelitian

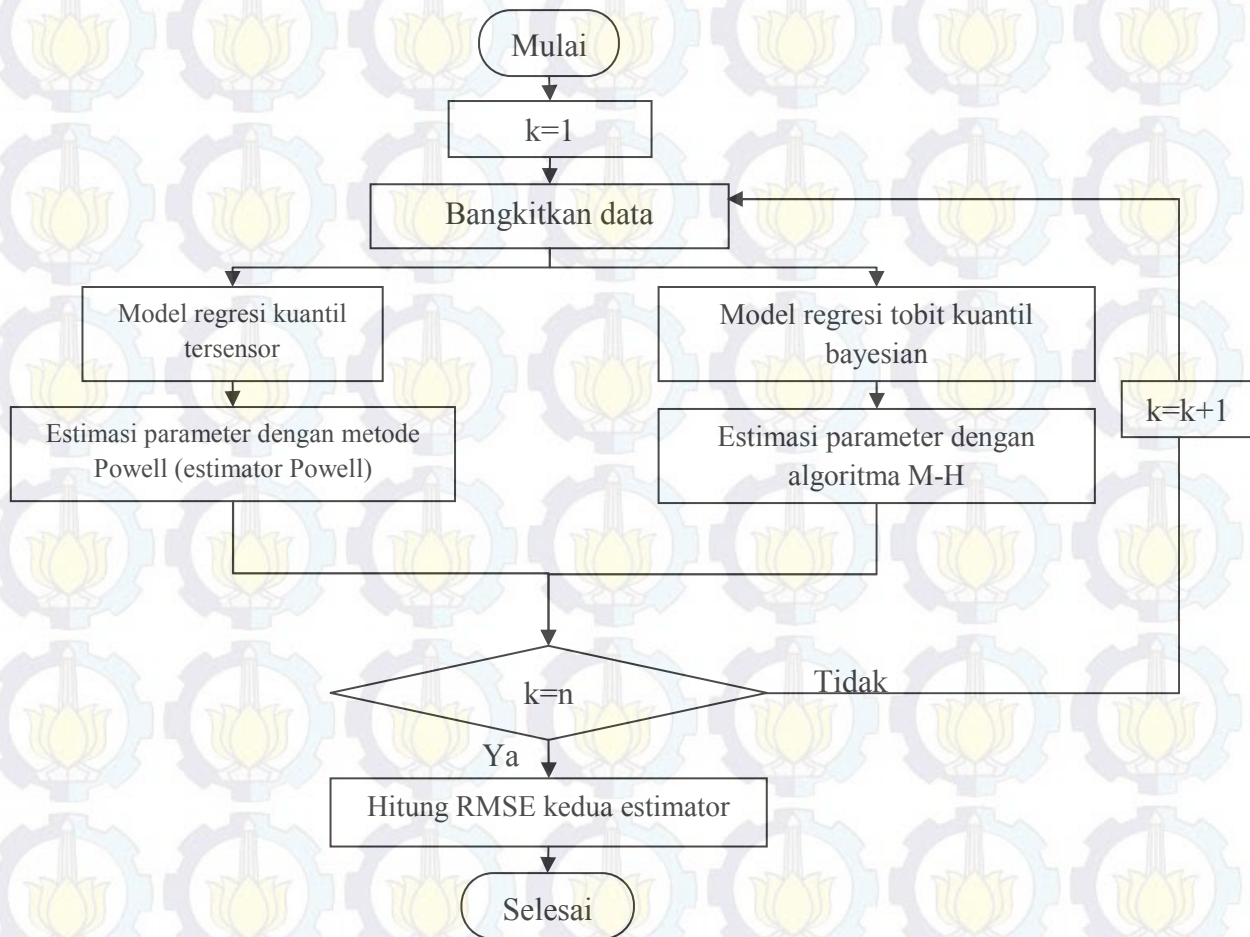
- Langkah-langkah untuk mengestimasi parameter model tobit kuantil bayesian dari distribusi posterior dan pemilihan model terbaik dengan pendekatan bayesian:



- a. Mensubstitusi model regresi tobit pada persamaan (2.5) ke dalam bentuk fungsi asimetris Laplace sehingga diperoleh fungsi asimetris Laplace tobit
  - b. Membentuk fungsi likelihood dari fungsi peluang asimetris Laplace tobit
  - c. Menghitung logaritma marginal likelihood dari model regresi tobit kuantil bayesian sesuai persamaan (2.23) untuk mendapatkan Bayes faktor.
2. Langkah-langkah untuk mengetahui performa estimator parameter model tobit kuantil bayesian dengan teknik simulasi yaitu membangkitkan data tersensor  $y_i$  dari model  $y_i = \max \left( 0.5 + \sum_{j=1}^p 0.5(j+1)(-1)^j x_j + \varepsilon_i, 0 \right)$  di mana  $p=1,3,5,7$  dengan tiga macam distribusi  $\varepsilon_i$  yakni (i) normal standar  $N(0,1)$ ; (ii) normal heteroskedastis  $(1+x_2)N(0,1)$ ; dan (iii) *mixture* normal  $0.75N(0,1)+0.25N(2,4)$ . Prosedur simulasi ini mengadaptasi penelitian dari Biliyas *et al.* (2000), Yu dan Stander (2007), dan Al-Hassan (2010). Proses perbandingan estimator regresi tobit kuantil bayesian dengan estimator Powell adalah sebagai berikut.
- a. Membangkitkan data tersensor  $y_i$  dengan beberapa kondisi yang dijabarkan sebelumnya di mana ukuran sampel  $n=100, 500, 1000$ , dan  $5000$
  - b. Untuk setiap pembangkitan data, dilakukan dua macam estimasi yaitu
    - i. Mengestimasi parameter model regresi tobit kuantil bayesian dengan teknik MCMC algoritma Metropolis-Hastings. Prosedur algoritma Metropolis-Hastings dapat dilihat pada Gambar 3.2.
    - ii. Mengestimasi parameter model regresi kuantil tersensor (estimator Powell) dengan metode BRCENS
  - c. Menghitung RMSE dari kedua estimator yang dihasilkan

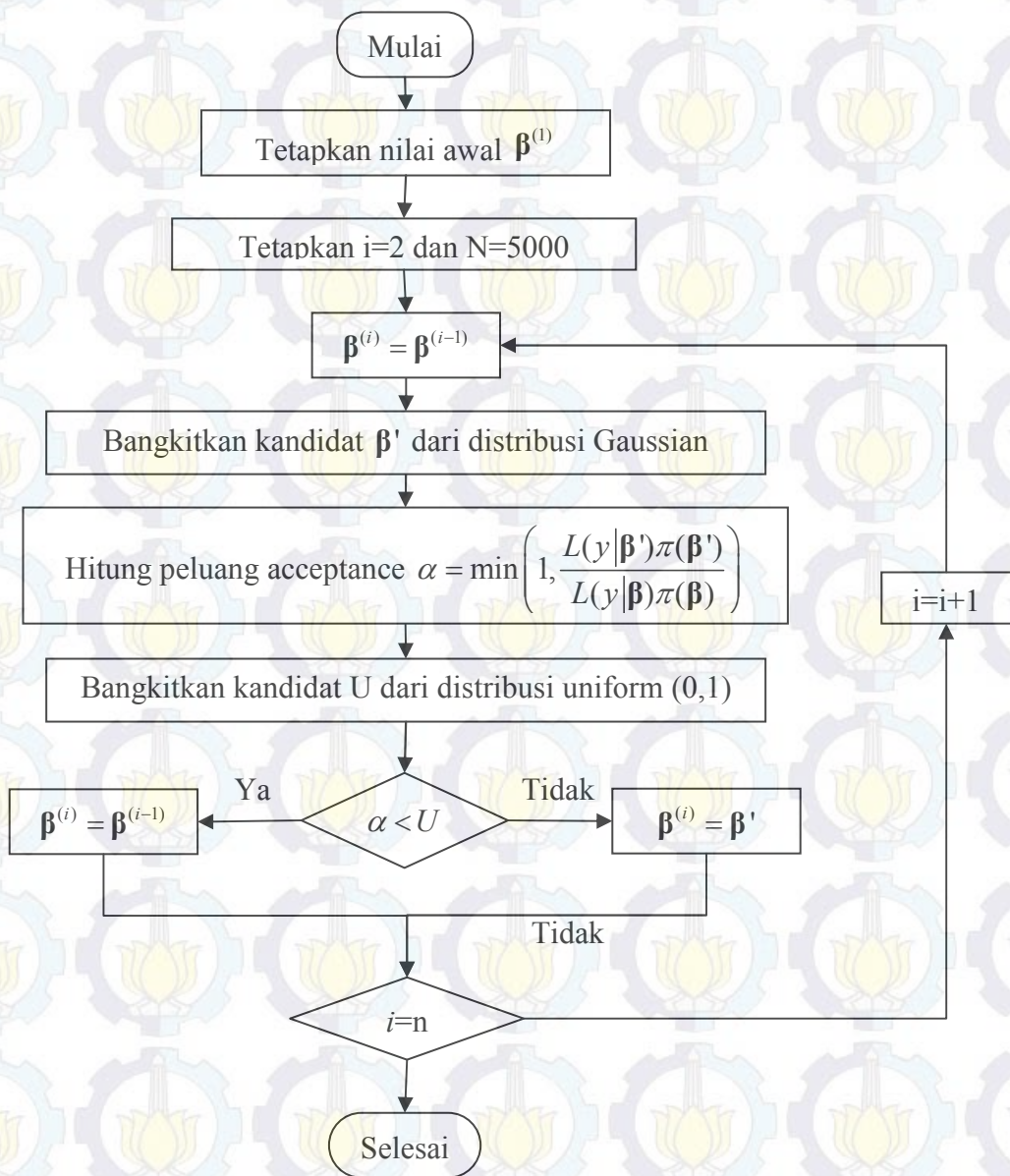
Untuk lebih jelas, alur perbandingan kedua estimator ini dapat dilihat dalam Gambar 3.1.





Gambar 3.1 Diagram Alir Algoritma Perbandingan Model Regresi Tobit Kuantil Bayesian dengan Model Regresi Kuantil Tersensor





Gambar 3.2 Diagram Alir Algoritma Metropolis-Hastings

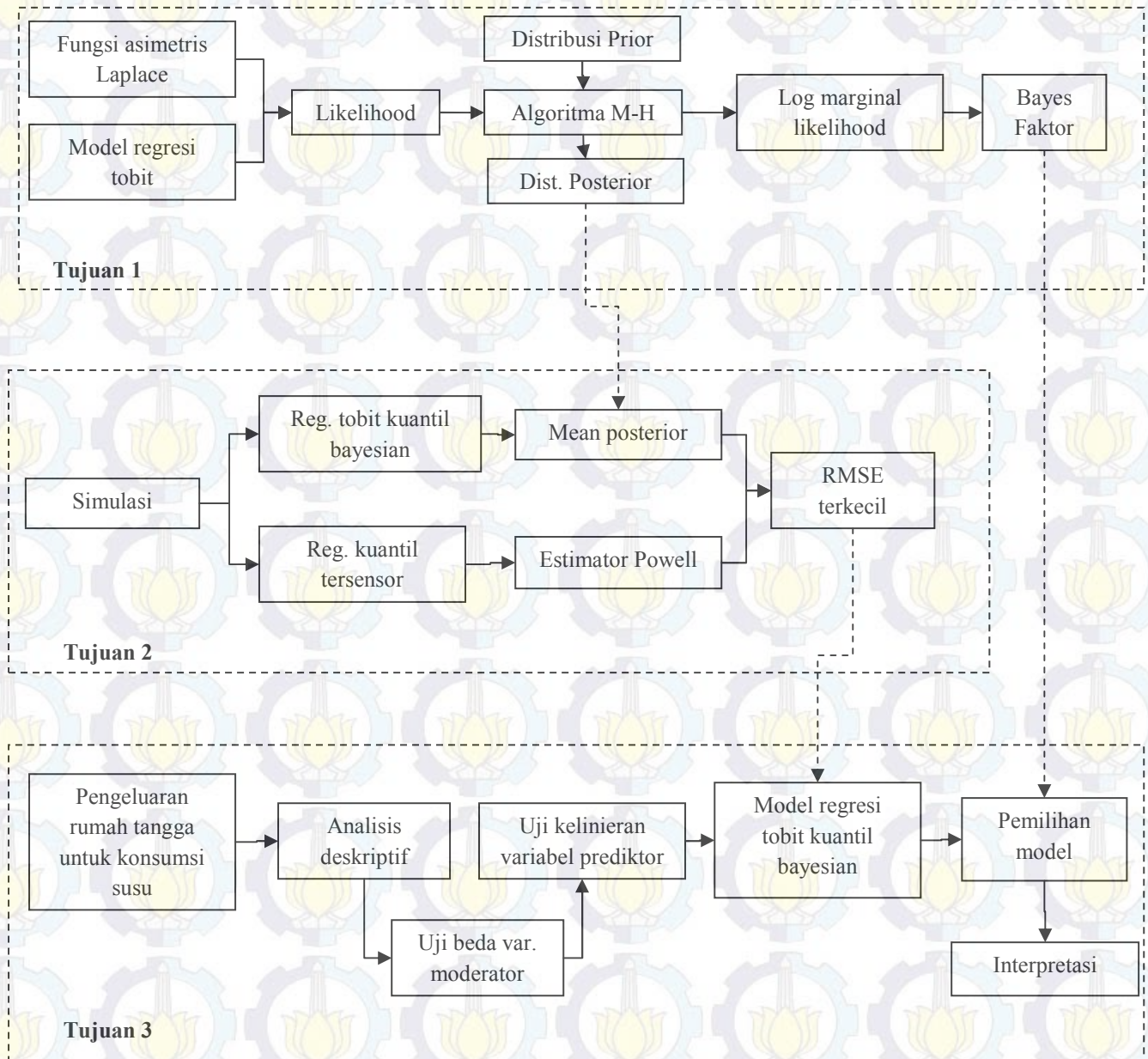


3. Langkah-langkah untuk menerapkan model regresi tobit kuantil bayesian terhadap pengeluaran rumah tangga untuk konsumsi susu dan mendapatkan model terbaiknya adalah sebagai berikut.

- a. Melakukan analisis deskriptif terhadap semua variabel respon dan prediktor yang digunakan dalam penelitian ini. Andaikan terdapat variabel yang memiliki variasi sangat besar, maka dilakukan transformasi.
- b. Melakukan uji beda antar wilayah (pedesaan dan perkotaan) untuk mengetahui peran variabel tersebut sebagai moderator.
- c. Melakukan uji hubungan linier antara masing-masing variabel prediktor terhadap variabel respon. Apabila terdapat prediktor yang nonlinier maka diusahakan transformasi agar menjadi linier. Namun jika tetap nonlinier, maka variabel prediktor tersebut tidak diikutsertakan dalam analisis selanjutnya
- d. Membentuk model regresi tobit kuantil untuk kuantil 0.05, 0.25, 0.50, 0.75 dan 0.95.
- e. Menentukan distribusi prior bagi parameter-paramater model regresi tobit kuantil bayesian pengeluaran rumah tangga untuk konsumsi susu
- f. Mendapatkan estimasi parameter model regresi tobit kuantil bayesian di mana estimator yang digunakan merupakan nilai rata-rata estimator yang dihasilkan dari algoritma Metropolis-Hastings
- g. Pemilihan model regresi tobit kuantil bayesian pengeluaran rumah tangga untuk konsumsi susu dengan kriteria Bayes faktor sesuai poin 1.d.
- h. Interpretasi hasil.

Ringkasan metode untuk ketiga tujuan dalam penelitian digambarkan dalam kerangka penelitian pada Gambar 3.3.





Gambar 3.3 Kerangka Penelitian



## BAB 4 HASIL DAN PEMBAHASAN

### 4.1 Estimasi Parameter dan Pemilihan Model Terbaik Regresi TKB

#### 4.1.1 Estimasi Parameter Model Regresi TKB

Berdasarkan persamaan (2.5), bentuk umum model regresi tobit sebagai berikut.

$$y_i = \max(\mathbf{x}_i' \boldsymbol{\beta} + \varepsilon_i, \tau)$$

Adapun persamaan berikut merupakan fungsi asimetris Laplace menurut (2.19).

$$f_{\theta}(u) = \theta(1-\theta) \exp(-\rho_{\theta}(u)) \quad , 0 < \theta < 1$$

Jika  $u = \varepsilon_i$  dan  $\tau = 0$ , maka  $u = y_i - \max(\mathbf{x}_i' \boldsymbol{\beta}, 0)$ . Oleh karena itu, persamaan (2.19) berubah menjadi (4.1).

$$f_{\theta}(\varepsilon_i) = \theta(1-\theta) \exp\{-\rho_{\theta}(y_i - \max(\mathbf{x}_i' \boldsymbol{\beta}, 0))\} \quad (4.1)$$

Setelah diketahui fungsi asimetris Laplace yang bersesuaian dengan model tobit, maka bisa didapatkan fungsi likelihood  $L(\mathbf{y} | \boldsymbol{\beta}(\theta))$  sebagai berikut.

$$\begin{aligned} L(\mathbf{y} | \boldsymbol{\beta}(\theta)) &= \prod_{i=1}^n f_{\theta}(\varepsilon_i) \\ &= \prod_{i=1}^n \theta(1-\theta) \exp\{-\rho_{\theta}(y_i - \max(\mathbf{x}_i' \boldsymbol{\beta}, 0))\} \\ &= \theta^n (1-\theta)^n \exp\left\{-\rho_{\theta} \sum_{i=1}^n (y_i - \max(\mathbf{x}_i' \boldsymbol{\beta}, 0))\right\} \end{aligned} \quad (4.2)$$

Distribusi posterior  $\pi(\boldsymbol{\beta}(\theta) | \mathbf{y})$  dapat dirumuskan sebagai (4.3).

$$\pi(\boldsymbol{\beta}(\theta) | \mathbf{y}) \propto L(\mathbf{y} | \boldsymbol{\beta}(\theta)) \pi(\boldsymbol{\beta}(\theta)) \quad (4.3)$$



Estimator model TKB diperoleh dengan menghitung nilai mean dari distribusi posterior parameter  $\pi(\boldsymbol{\beta}(\theta)|\mathbf{y})$ .

#### 4.1.2 Pemilihan Model Terbaik Regresi TKB

Sebagaimana telah disebutkan pada bab sebelumnya, distribusi posetrior  $\pi(\boldsymbol{\beta}(\theta)|\mathbf{y})$  diperoleh dengan algoritma Metropolis-Hastings. Dari algoritma ini dapat dihitung nilai marginal likelihood sebagai komponen Bayes faktor yang digunakan untuk melakukan pemilihan model terbaik. Marginal likelihood dari model  $M_k$  didefinisikan sebagai berikut.

$$m(\mathbf{y}|M_k) = \int L(\mathbf{y}|M_k, \boldsymbol{\beta}(\theta)_k) \pi(\boldsymbol{\beta}(\theta)_k | M_k) d\boldsymbol{\beta}(\theta)_k \quad (4.4)$$

Untuk selanjutnya, penulisan  $\boldsymbol{\beta}(\theta)$  disederhanakan menjadi  $\boldsymbol{\beta}$ .

Persamaan (4.4) adalah *normalizing constant* distribusi posterior  $\pi(\boldsymbol{\beta}_k|\mathbf{y})$ , sehingga persamaan ini dapat ditulis ulang sebagai berikut.

$$m(\mathbf{y}|M_k) = \frac{L(\mathbf{y}|M_k, \boldsymbol{\beta}_k) \pi(\boldsymbol{\beta}_k | M_k)}{\pi(\boldsymbol{\beta}_k | \mathbf{y}, M_k)} \quad (4.5)$$

Marginal likelihood dalam (4.5) dapat diestimasi dengan menghitung ordinat posterior  $\pi(\boldsymbol{\beta}_k^*|\mathbf{y}, M_k)$ , di mana elemen  $\boldsymbol{\beta}_k^*$  merupakan nilai-nilai yang memiliki densitas tinggi pada distribusi posterior (misal: mean). Apabila persamaan (4.5) ditambahkan logaritma dan disubstitusi  $\boldsymbol{\beta}_k = \boldsymbol{\beta}_k^*$ , maka diperoleh persamaan (4.6).

$$\log \hat{m}(\mathbf{y}|M_k) = \log L(\mathbf{y}|M_k, \boldsymbol{\beta}_k^*) + \log \pi(\boldsymbol{\beta}_k^* | M_k) - \log \pi(\boldsymbol{\beta}_k^* | \mathbf{y}, M_k) \quad (4.6)$$

Substitusi (4.2) ke (4.6) menghasilkan persamaan (4.7).

$$\begin{aligned} \log \hat{m}(\mathbf{y}|M_k) = & n \log \{ \theta(1-\theta) \} - \left\{ \rho_\theta \sum_{i=1}^n (y_i - maks(\mathbf{x}_i' \boldsymbol{\beta}_k^*, 0)) \right\} \\ & + \log \pi(\boldsymbol{\beta}_k^* | M_k) - \log \pi(\boldsymbol{\beta}_k^* | \mathbf{y}, M_k) \end{aligned} \quad (4.7)$$



Menurut Chib dan Jeliakoz (2001), ordinat posterior  $\pi(\beta^* | y)$  dapat diestimasi dengan persamaan (4.8).

$$\pi(\beta_k^* | y) = \frac{G^{-1} \sum_{i=1}^G \alpha(\beta^{(g)}, \beta^*) \pi(\beta^{(g)}, \beta^*)}{J^{-1} \sum_{i=1}^J \alpha(\beta^*, \beta^{(j)})} \quad (4.8)$$

di mana:

$$\begin{aligned} \alpha(\beta, \beta^*) &= \min \left\{ 1, \frac{\pi(\beta^*)}{\pi(\beta)} L^*(\beta^*, \beta) \right\} \\ \{\beta^{(j)}\} &= \text{sampel yang berasal dari distribusi proposal } \pi(\beta^*, \beta) \\ \{\beta^{(g)}\} &= \text{sampel yang berasal dari distribusi posterior} \end{aligned}$$

Substitusi persamaan (4.2) ke  $L^*(\beta^*, \beta)$  menghasilkan (4.9).

$$\begin{aligned} L^*(\beta^*, \beta) &= \frac{L(\beta^*)}{L(\beta)} \\ &= \frac{\theta^n (1-\theta)^n \exp \left\{ -\rho_\theta \sum_{i=1}^n (y_i - \text{maks}(\mathbf{x}_i' \beta^*, 0)) \right\}}{\theta^n (1-\theta)^n \exp \left\{ -\rho_\theta \sum_{i=1}^n (y_i - \text{maks}(\mathbf{x}_i' \beta, 0)) \right\}} \\ &= \exp \left\{ -\sum_{i=1}^n \left[ \rho_\theta (y_i - \text{maks}(\mathbf{x}_i' \beta^*, 0)) - \rho_\theta (y_i - \text{maks}(\mathbf{x}_i' \beta, 0)) \right] \right\} \quad (4.9) \end{aligned}$$

Ketika ordinat posterior berhasil diestimasi, Bayes faktor dari dua model, misal  $M_1$  dan  $M_2$ , dapat dihitung seperti dalam persamaan (4.10).

$$B_{12} = \exp \left\{ \log \hat{m}(y | M_1) - \log \hat{m}(y | M_2) \right\} \quad (4.10)$$

## 4.2 Perbandingan Performa Estimator TKB dan Powell

Ada beberapa aspek yang diperhatikan dalam membandingkan estimator regresi TKB dengan estimator Powell yakni menurut banyaknya prediktor, ukuran sampel dan distribusi error/residual yang berbeda. Ukuran pembanding yang

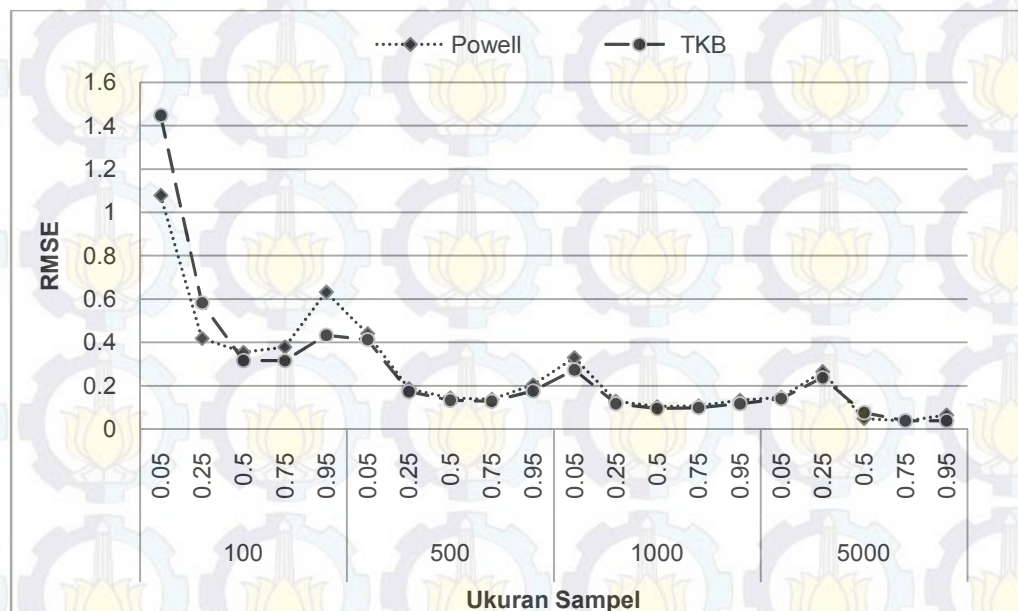


digunakan adalah RMSE, di mana semakin kecil RMSE maka semakin baik performa suatu estimator. Perbandingan pertama menyangkut banyaknya prediktor dalam model tobit ditampilkan dalam Tabel 4.1.

Tabel 4.1 RMSE Intersep ( $\beta_0$ ) dari Estimator TKB dan Powell dengan Banyaknya Prediktor Berbeda (n=100)

| Banyaknya Prediktor | TKB   |       |       |       |       | Powell |       |       |       |       |
|---------------------|-------|-------|-------|-------|-------|--------|-------|-------|-------|-------|
|                     | 0.05  | 0.25  | 0.50  | 0.75  | 0.95  | 0.05   | 0.25  | 0.50  | 0.75  | 0.95  |
| 1                   | 1.857 | 1.008 | 0.274 | 0.154 | 0.195 | 1.288  | 0.927 | 0.208 | 0.151 | 0.215 |
| 3                   | 1.448 | 0.583 | 0.316 | 0.316 | 0.433 | 1.079  | 0.418 | 0.355 | 0.378 | 0.632 |
| 5                   | 2.931 | 0.560 | 0.480 | 0.429 | 0.597 | 0.869  | 0.530 | 0.404 | 0.440 | 0.998 |
| 7                   | 1.362 | 0.629 | 0.458 | 0.416 | 0.701 | 1.264  | 0.652 | 0.500 | 0.541 | 1.766 |

Tabel 4.1 hanya menampilkan besarnya RMSE dari estimator intersep model tobit dengan jumlah prediktor yang berbeda. Berdasarkan tabel ini dapat dilihat bahwa secara umum performa estimator TKB lebih baik daripada Powell. Selain itu, dapat pula diketahui bahwa semakin banyak prediktor yang digunakan, maka RMSE cenderung semakin besar. Atau dengan kata lain, performa estimator semakin menurun.



Gambar 4.1 RMSE Intersep ( $\beta_0$ ) dari Estimator TKB dan Powell dengan Jumlah Sampel Berbeda (p=3)



Gambar 4.1 memperlihatkan perubahan RMSE estimator TKB dan Powell seiring peningkatan ukuran sampel untuk intersep. Pada kuantil 0.05 dengan ukuran sampel 100-500, RMSE dari kedua estimator bernilai cukup besar, terutama estimator TKB. Di sisi lain, ketika ukuran sampel 1000 atau lebih, nilai RMSE menurun drastis. Dalam kondisi ini, RMSE estimator TKB lebih kecil daripada estimator Powell. Hal yang sama juga berlaku untuk parameter lain, yang dapat dilihat pada Lampiran 2.

Berdasarkan Tabel 4.1 dan Gambar 4.1, dapat disimpulkan bahwa untuk ukuran sampel kecil dan kuantil bawah (0.05), performa estimator TKB dan Powell tidak cukup baik. Namun, ketika ukuran sampel cukup besar ( $>1000$ ), kedua estimator ini memiliki performa yang bagus. Selain itu, pada kondisi ini RMSE estimator TKB lebih kecil daripada Powell, atau dengan kata lain performa estimator TKB lebih baik daripada Powell. Di sisi lain, melihat adanya kecenderungan RMSE yang selalu bernilai besar untuk model kuantil bawah ( $\theta=0.05$ ), sehingga dapat dikatakan bahwa baik estimator TKB maupun Powell tidak cukup baik untuk mengestimasi model pada kuantil bawah. Selanjutnya, dengan menggunakan banyaknya prediktor dan jumlah sampel yang cukup optimal ( $n=1000$  dan  $p=3$ ) bagi estimator TKB dan Powell, dilakukan simulasi untuk membandingkan performa estimator-estimator tersebut pada kondisi distribusi error yang berbeda yaitu normal, *heteroscedastics*, dan *mixture*. Pemilihan distribusi error mengadopsi penelitian Yu dan Stander (2007).

Tabel 4.2 Perbandingan RMSE dari Model Tobit dengan Distribusi Error Berbeda ( $n=1000$ ,  $\theta=0.50$ )

| Estimator | Normal |        |       | <i>Heteroscedastics</i> |        |       | <i>Mixture</i> |        |       |
|-----------|--------|--------|-------|-------------------------|--------|-------|----------------|--------|-------|
|           | TKB    | Powell | Tobit | TKB                     | Powell | Tobit | TKB            | Powell | Tobit |
| $\beta_0$ | 0.094  | 0.105  | 0.070 | 0.115                   | 0.125  | 0.181 | 0.354          | 0.355  | 0.311 |
| $\beta_1$ | 0.050  | 0.053  | 0.042 | 0.064                   | 0.067  | 0.059 | 0.055          | 0.060  | 0.133 |
| $\beta_2$ | 0.146  | 0.154  | 0.113 | 0.220                   | 0.235  | 0.311 | 0.229          | 0.237  | 0.349 |
| $\beta_3$ | 0.066  | 0.065  | 0.045 | 0.084                   | 0.087  | 0.066 | 0.075          | 0.077  | 0.222 |

Dari Tabel 4.2 dapat diketahui bahwa secara umum RMSE estimator TKB lebih kecil daripada Powell untuk ketiga distribusi error yang digunakan. Hal ini menunjukkan bahwa performa estimator TKB lebih baik daripada Powell. Di



samping itu, Tabel 4.2 juga mengikutsertakan estimator yang berasal dari regresi tobit standar untuk dibandingkan dengan estimator TKB dan Powell. Agar ketiga estimator tersebut layak diperbandingkan, maka kuantil yang digunakan untuk estimator TKB dan Powell adalah 0.50, di mana kuantil ini setara dengan nilai mean yang merupakan dasar estimasi parameter regresi tobit standar (*conditional mean*). Hasil simulasi menunjukkan bahwa performa estimator TKB dan Powell tidak lebih baik dari estimator tobit standar untuk model dengan distribusi error normal. Namun, untuk dua distribusi error lain yaitu heteroskedastis dan normal, performa estimator TKB dan Powell lebih baik daripada estimator tobit standar, karena memiliki nilai RMSE lebih kecil.

#### 4.3 Pemodelan Regresi TKB Bagi Pengeluaran Rumah Tangga untuk Konsumsi Susu

##### 4.3.1 Deskripsi Data Penelitian

Statistik deskriptif dari data yang digunakan dalam penelitian ini (SUSETI) ditampilkan dalam Tabel 4.3 dan Tabel 4.4 sebagai berikut.

Tabel 4.3 Statistik Deskriptif Pengeluaran Rumah Tangga untuk Konsumsi Menurut Wilayah

| Statistik          | Wilayah   | Pengeluaran RT<br>untuk Konsumsi susu<br>(ribuan rupiah) |
|--------------------|-----------|--|
| Minimum            | Pedesaan  | 0.00   |
|                    | Perkotaan | 0.00   |
| Maksimum           | Pedesaan  | 458.000  |
|                    | Perkotaan | 750.000  |
| Rata-rata          | Pedesaan  | 8.868  |
|                    | Perkotaan | 19.341   |
| Standar<br>Deviasi | Pedesaan  | 23.378   |
|                    | Perkotaan | 39.786   |

Berdasarkan Tabel 4.3 dapat diketahui bahwa baik di wilayah pedesaan dan perkotaan terdapat rumah tangga yang sama sekali tidak mengeluarkan biaya



untuk konsumsi susu dengan persentase masing-masing 54.68% dan 36.61%. Hal ini menunjukkan bahwa rumah tangga di perkotaan lebih banyak yang memiliki pengeluaran untuk konsumsi susu dibandingkan di pedesaan, sehingga tidak mengherankan jika nilai maksimum dan rata-rata pengeluaran untuk konsumsi susu di perkotaan juga lebih tinggi.

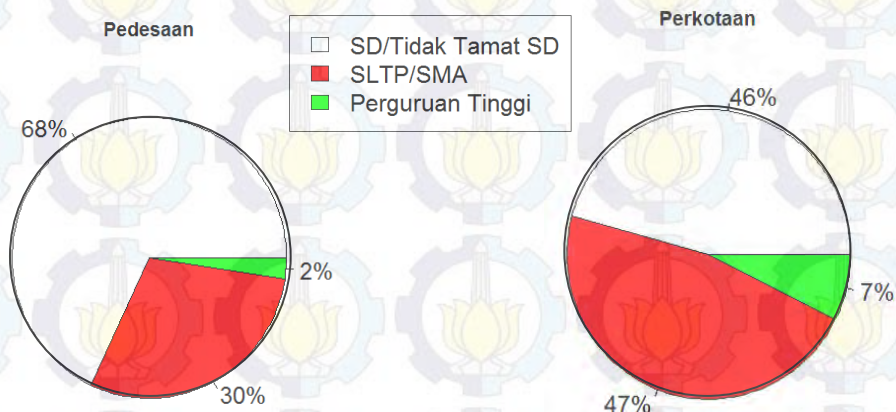
Tabel 4.4 Statistik Deskriptif Variabel Prediktor Penelitian Bersifat Kontinu

| Statistik       | Variabel Wilayah | X <sub>1</sub> | X <sub>3</sub> | X <sub>4</sub> | X <sub>5</sub> | X <sub>6</sub> | X <sub>7</sub> |
|-----------------|------------------|----------------|----------------|----------------|----------------|----------------|----------------|
| Minimum         | Pedesaan         | 82.00          | 1.12           | 1.00           | 10.00          | 0.00           | 10.42          |
|                 | Perkotaan        | 100.00         | 0.69           | 2.00           | 8.33           | 0.00           | 15.74          |
| Maksimum        | Pedesaan         | 80000.00       | 92.66          | 18.00          | 100.00         | 75.00          | 4296.15        |
|                 | Perkotaan        | 34000.00       | 91.52          | 19.00          | 100.00         | 75.00          | 8416.03        |
| Rata-rata       | Pedesaan         | 1076.81        | 50.62          | 4.85           | 41.72          | 25.56          | 170.26         |
|                 | Perkotaan        | 1780.51        | 44.65          | 5.45           | 37.80          | 25.48          | 233.44         |
| Standar Deviasi | Pedesaan         | 1888.94        | 19.73          | 1.61           | 17.52          | 15.81          | 286.26         |
|                 | Perkotaan        | 2184.92        | 19.02          | 1.95           | 16.49          | 15.95          | 439.93         |

Berdasarkan Tabel 4.4 dapat diketahui bahwa rata-rata dan standar deviasi pendapatan rumah tangga yang ada di wilayah pedesaan lebih rendah daripada wilayah perkotaan. Padahal, jarak antar nilai minimum dan maksimum pendapatan di pedesaan lebih lebar. Selain itu, rata-rata persentase pengeluaran pangan rumah tangga di pedesaan lebih besar daripada perkotaan, sedangkan rata-rata dan standar deviasi pengeluaran per kapita rumah tangga di perkotaan lebih besar daripada pedesaan.

Menurut aspek demografis, rata-rata jumlah anggota rumah tangga dan persentase anggota rumah tangga yang bekerja di pedesaan lebih besar daripada wilayah kota, sedangkan rata-rata persentase anggota rumah tangga yang berusia <12 tahun di pedesaan memiliki karakteristik yang hampir serupa dengan di perkotaan.





Gambar 4.2 Diagram Pie Pendidikan Kepala Rumah Tangga Menurut Wilayah

Di sisi lain, dari Gambar 4.2 dapat diketahui bahwa baik wilayah pedesaan maupun perkotaan didominasi oleh kepala rumah tangga yang berpendidikan SD atau tidak tamat SD. Namun, persentase kepala rumah tangga yang menempuh pendidikan menengah dan tinggi di perkotaan lebih besar daripada di wilayah pedesaan.

#### 4.3.2 Hasil Uji Beda Karakteristik Wilayah

Uji beda terhadap karakteristik wilayah pedesaan dan perkotaan menggunakan uji t dan uji independensi *chi-square*. Adapun yang dimaksud dengan karakteristik adalah variabel-variabel lain kecuali variabel wilayah yang termasuk dalam penelitian ini. Uji t digunakan untuk karakteristik yang bersifat kontinu, sedangkan uji *chi-square* untuk kategorik. Uji-uji ini bertujuan untuk mengetahui apakah variabel wilayah mampu berperan sebagai variabel moderator dalam analisis selanjutnya. Hasil pengujian beda ini ditampilkan dalam Tabel 4.5 dan Tabel 4.6.



Tabel 4.5 Hasil Uji t Karakteristik Wilayah Pedesaan dan Perkotaan

| Variabel                                | Statistik uji t | P-value | Keterangan    |
|---|-----------------|---------|---------------|
| Pendapatan rumah tangga ( $X_1$ )       | -11.959         | 0.000   | Beda          |
| Persentase pengeluaran pangan ( $X_3$ ) | 10.976          | 0.000   | Beda          |
| Jumlah anggota rumah tangga ( $X_4$ )   | -11.683         | 0.000   | Beda          |
| Persentase ART bekerja ( $X_5$ )        | 8.234           | 0.000   | Beda          |
| Persentase ART usia <12 tahun ( $X_6$ ) | 0.177           | 0.860   | Tidak berbeda |
| Pengeluaran per kapita ( $X_7$ )        | -5.714          | 0.000   | Beda          |
| Konsumsi susu (Y)                       | -10.663         | 0.000   | Beda          |

Tabel 4.6 Hasil Uji *chi-square* Karakteristik Wilayah Pedesaan dan Perkotaan

| Variabel                                | Statistik uji <i>chi-square</i> | P-value | Keterangan |
|---|---------------------------------|---------|------------|
| Pendidikan menengah-rendah ( $X_{21}$ ) | 163.656                         | 0.000   | Beda       |
| Pendidikan atas-rendah ( $X_{22}$ )     | 73.186                          | 0.000   | Beda       |

Berdasarkan Tabel 4.5 dan Tabel 4.6 dapat diketahui bahwa hampir semua karakteristik/variabel antar wilayah pedesaan dan perkotaan menunjukkan perbedaan, kecuali persentase ART usia <12 tahun. Oleh karena itu, dapat dikatakan bahwa variabel wilayah dapat dijadikan sebagai variabel moderator dalam analisis lebih lanjut.

#### 4.3.3 Hasil Pengujian Linieritas Variabel Prediktor

Hasil uji linieritas variabel prediktor kontinu terhadap variabel respon menggunakan uji RESET dapat dilihat pada Tabel 4.7, sedangkan uji linieritas untuk prediktor dummy menggunakan uji korelasi *biserial point* disajikan dalam Tabel 4.8.



Tabel 4.7 Hasil Uji RESET

| Variabel Prediktor                      | Wilayah   | Statistik Uji | <i>P-value</i> | Keterangan |
|---|-----------|---------------|----------------|------------|
| Pendapatan rumah tangga ( $X_1$ )       | Pedesaan  | 7.077         | 0.008          | Nonlinier  |
|   | Perkotaan | 48.131        | 0.000          | Nonlinier  |
| Persentase pengeluaran pangan ( $X_3$ ) | Pedesaan  | 0.003         | 0.958          | Linier     |
|   | Perkotaan | 3.683         | 0.055          | Linier     |
| Jumlah anggota rumah tangga ( $X_4$ )   | Pedesaan  | 0.769         | 0.381          | Linier     |
|   | Perkotaan | 0.000         | 0.639          | Linier     |
| Persentase ART bekerja ( $X_5$ )        | Pedesaan  | 0.000         | 0.983          | Linier     |
|   | Perkotaan | 0.644         | 0.422          | Linier     |
| Persentase ART <12 tahun ( $X_6$ )      | Pedesaan  | 3.558         | 0.059          | Linier     |
|   | Perkotaan | 0.644         | 0.422          | Linier     |
| Pengeluaran per kapita ( $X_7$ )        | Pedesaan  | 6.201         | 0.013          | Nonlinier  |
|   | Perkotaan | 24.397        | 0.000          | Nonlinier  |

Berdasarkan Tabel 4.7 dapat diketahui bahwa sebagian besar variabel prediktor yang bersifat kontinu berhubungan linier dengan variabel respon konsumsi susu. Kecuali prediktor pendapatan rumah tangga dan pengeluaran per kapita untuk wilayah perkotaan dan pedesaan. Agar prediktor-prediktor tersebut bisa tetap dipertahankan, maka dilakukan beberapa transformasi yang tidak mengubah *magnitude* data yaitu logaritma, kuadrat, dan akar kuadrat. Dari upaya-upaya ini, prediktor yang berhasil dilinierkan adalah pengeluaran per kapita ( $X_7$ ) untuk pedesaan dan pendapatan rumah tangga ( $X_1$ ) untuk perkotaan melalui transformasi logaritma dengan nilai *p-value* masing-masing adalah 0.317 dan 0.151. Hasil pengujian linieritas dengan transformasi selengkapnya dapat dilihat pada Lampiran 5. Variabel-variabel prediktor yang bersifat nonlinier terhadap respon tidak diikutsertakan dalam analisis lebih lanjut.

Tabel 4.8 Hasil Uji Korelasi *Point Biserial* Prediktor Dummy  $X_2$ 

| Variabel Prediktor                      | Wilayah   | Korelasi <i>Biserial</i> | Statistik Uji <i>t</i> | <i>P-value</i> | Keterangan |
|---|-----------|--------------------------|------------------------|----------------|------------|
| Pendidikan menengah-rendah ( $X_{21}$ ) | Pedesaan  | -0.096                   | -3.824                 | 0.000          | Linier     |
|   | Perkotaan | -0.022                   | -0.758                 | 0.224          | Nonlinier  |
| Pendidikan tinggi-rendah ( $X_{22}$ )   | Pedesaan  | -0.178                   | -7.186                 | 0.000          | Linier     |
|   | Perkotaan | -0.191                   | -6.847                 | 0.000          | Linier     |



Tabel 4.8 memperlihatkan bahwa variabel *dummy* untuk pendidikan kepala rumah tangga di wilayah pedesaan menunjukkan hubungan yang linier dengan variabel respon. Di sisi lain, untuk wilayah perkotaan, hasil uji korelasi *Pearson* untuk *dummy*  $X_{22}$  mendukung adanya hubungan linier dengan respon, sedangkan *dummy*  $X_{21}$  tidak. Namun demikian, karena variabel-variabel *dummy* ini pada dasarnya berasal dari satu variabel, maka keduanya tetap disertakan dalam pemodelan regresi TKB.

#### 4.3.4 Hasil Estimasi Parameter Regresi TKB

Sampel yang digunakan sebagai data penelitian ini berasal dari SUSETI dengan jumlah sebanyak 3,471 unit untuk pedesaan dan 1,961 unit untuk perkotaan. Jumlah-jumlah sampel tersebut sangat sesuai untuk pemodelan regresi TKB, sebagaimana hasil simulasi yang telah dijabarkan pada subbab 4.2 sebelumnya. Estimasi parameter regresi TKB dilakukan untuk lima titik kuantil yaitu 0.05, 0.25, 0.50, 0.75 dan 0.95. Adapun hasil estimasi ini disajikan dalam Tabel 4.8 dan 4.9.

Tabel 4.9 Hasil Estimasi Parameter Regresi TKB Wilayah Pedesaan

| Parameter    | Estimator TKB |         |        |        |        |
|--------------|---------------|---------|--------|--------|--------|
|              | Kuantil       |         |        |        |        |
|              | 0.05          | 0.25    | 0.50   | 0.75   | 0.95   |
| $\beta_0$    | -0.464        | 0.8988  | -0.741 | 0.275  | 0.735  |
| $\beta_{21}$ | -2.633        | 0.568   | 0.400  | 0.862  | 3.056  |
| $\beta_{22}$ | 1.288         | -4.3081 | 0.460  | 0.594  | 0.775  |
| $\beta_3$    | -4.103        | -2.2482 | -0.034 | 0.001  | -0.044 |
| $\beta_4$    | 0.374         | -0.3244 | 0.250  | 0.554  | 4.430  |
| $\beta_5$    | -2.479        | -4.5692 | -0.164 | -0.054 | -0.283 |
| $\beta_6$    | -5.281        | -2.7559 | -0.068 | -0.076 | -0.335 |
| $\beta_7$    | 0.082         | 2.7235  | 1.860  | 2.129  | 8.068  |

Dari Tabel 4.9 dapat diketahui bahwa estimator TKB untuk pengeluaran konsumsi susu di pedesaan bervariasi antar kuantil. Hal ini mengindikasikan bahwa penggunaan model regresi TKB cukup tepat untuk memodelkan data tersebut. Besarnya estimator parameter untuk variabel pendidikan kepala rumah tangga menengah-rendah ( $X_{21}$ ), persentase ART bekerja ( $X_5$ ), persentase ART



usia <12 tahun ( $X_6$ ), dan rata-rata pengeluaran per kapita ( $X_7$ ) cenderung semakin meningkat seiring pertambahan nilai kuantil. Keadaan ini menunjukkan bahwa pengaruh variabel-variabel tersebut semakin besar untuk pengeluaran konsumsi susu rumah tangga di pedesaan yang semakin tinggi. Di sisi lain, estimator parameter untuk variabel pendidikan kepala rumah tangga tinggi-rendah ( $X_{22}$ ), persentase pengeluaran pangan ( $X_3$ ), dan jumlah anggota rumah tangga ( $X_4$ ), bersifat fluktuatif antar kuantil.

Tabel 4.10 Hasil Estimasi Parameter Regresi TKB Wilayah Perkotaan

| Parameter    | Estimator TKB |        |        |        |        |
|--------------|---------------|--------|--------|--------|--------|
|              | Kuantil       |        |        |        |        |
|              | 0.05          | 0.25   | 0.50   | 0.75   | 0.95   |
| $\beta_0$    | -4.229        | -0.043 | 0.624  | -0.363 | -1.021 |
| $\beta_1$    | -1.989        | 2.929  | 1.663  | 4.351  | 12.829 |
| $\beta_{21}$ | 1.012         | 0.206  | 0.606  | 0.545  | 1.910  |
| $\beta_{22}$ | 3.056         | 0.941  | 0.280  | 1.575  | 3.385  |
| $\beta_3$    | -5.158        | -0.041 | -0.062 | -0.138 | -0.388 |
| $\beta_4$    | 2.668         | 0.481  | 0.101  | 1.818  | 4.854  |
| $\beta_5$    | -4.562        | -1.341 | -0.056 | -0.176 | -0.413 |
| $\beta_6$    | -0.148        | -0.068 | -0.054 | -0.213 | -0.343 |

Berdasarkan Tabel 4.10 dapat diketahui bahwa perilaku estimator TKB di wilayah perkotaan mirip dengan wilayah pedesaan, dalam artian bahwa estimator-estimator tersebut bervariasi antar kuantil. Estimator parameter untuk variabel pendapatan rumah tangga ( $X_1$ ) cenderung semakin meningkat seiring pertambahan nilai kuantil. Keadaan ini menunjukkan bahwa pengaruh variabel tersebut semakin besar untuk pengeluaran konsumsi susu rumah tangga di pedesaan yang semakin tinggi. Selain itu, estimator untuk variabel pendidikan kepala rumah tangga ( $X_{21}$  dan  $X_{22}$ ), persentase pengeluaran pangan ( $X_3$ ), jumlah anggota rumah tangga ( $X_4$ ), persentase ART bekerja ( $X_5$ ) dan persentase ART usian <12 tahun ( $X_6$ ) menunjukkan bahwa pengaruh variabel-variabel tersebut berubah secara fluktuatif terhadap perubahan nilai kuantil.

#### 4.3.5 Pemilihan Model Regresi TKB Terbaik

Pemilihan model terbaik dalam penelitian ini bertujuan untuk mengetahui variabel prediktor mana saja yang berperan dalam pembentukan model regresi



TKB dengan menggunakan Bayes faktor, di mana kriteria pemilihan yang dipakai berdasarkan Tabel 2.1. Hasil perhitungan Bayes faktor untuk model TKB wilayah pedesaan dan perkotaan ditunjukkan oleh Tabel 4.11 dan 4.12, di mana  $M_1$  merupakan model penuh yaitu model dengan prediktor yang berhubungan linier dengan respon.

Tabel 4.11 Bayes Faktor (2 Log B) Model Regresi TKB Reduksi Pedesaan

| Model Reduksi ( $M_2$ )                 | Kuantil             |                     |                        |                        |                         |
|---|---------------------|---------------------|------------------------|------------------------|-------------------------|
|   | 0.05                | 0.25                | 0.50                   | 0.75                   | 0.95                    |
| Pend. Menengah-rendah ( $X_{21}$ )      | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 136.39 <sup>(**)</sup> | 753.67 <sup>(**)</sup> | 821.6 <sup>(**)</sup>   |
| Pend. Tinggi-rendah ( $X_{22}$ )        | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 14.29 <sup>(**)</sup>  | 192.49 <sup>(**)</sup> | 49.21 <sup>(**)</sup>   |
| Persentase Pengeluaran pangan ( $X_3$ ) | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | -71.84 <sup>(-)</sup>  | -26.33 <sup>(-)</sup>  | -180.24 <sup>(-)</sup>  |
| Jumlah anggota rumah tangga ( $X_4$ )   | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 222.32 <sup>(**)</sup> | 307.48 <sup>(**)</sup> | 521.86 <sup>(**)</sup>  |
| Persentase ART bekerja ( $X_5$ )        | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 90.42 <sup>(**)</sup>  | 52.03 <sup>(**)</sup>  | 144.32 <sup>(**)</sup>  |
| Persentase ART usia <12 tahun ( $X_6$ ) | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 28.47 <sup>(**)</sup>  | 231.29 <sup>(**)</sup> | 869.97 <sup>(**)</sup>  |
| Pengeluaran per kapita ( $X_7$ )        | 0.01 <sup>(0)</sup> | 0.01 <sup>(0)</sup> | 209.27 <sup>(**)</sup> | 830.28 <sup>(**)</sup> | 1042.66 <sup>(**)</sup> |

Keterangan: Kekuatan pembuktian terhadap  $M_1$ , (-) : negatif, (0) : tidak ada; (+) : positif, (\*) : kuat; (\*\*): sangat kuat

Berdasarkan Tabel 4.11 dapat diketahui bahwa model terbaik kuantil 0.50, 0.75, dan 0.95 adalah model yang tidak mengikutsertakan prediktor persentase pengeluaran pangan ( $X_3$ ). Adapun model terbaik kuantil 0.05 dan 0.25 merupakan model penuh. Hal yang menarik dari hasil ini adalah besarnya Bayes faktor untuk model pada kuantil 0.05 dan 0.25 yang menunjukkan nilai mendekati nol, atau dengan kata lain Bayes faktor tidak dapat memutuskan mana di antara model penuh dan model reduksi yang sesuai untuk memodelkan variabel respon. Namun demikian, dalam penelitian ini diputuskan bahwa apabila terdapat perbandingan dua model yang menghasilkan Bayes faktor mendekati nol, maka model yang digunakan adalah model penuh. Hal ini dikarenakan jika yang dipakai adalah model reduksi, maka dikhawatirkan akan berujung pada terpilihnya model konstan yaitu model tanpa prediktor.



Model-model terbaik regresi TKB bagi pengeluaran rumah tangga untuk konsumsi susu di pedesaan dapat dituliskan sebagai persamaan-persamaan berikut.

$$\hat{y}_{(0.05)} = \begin{cases} -0.464 - 2.633X_{21} + 1.288X_{22} - 4.103X_3 + 0.374X_4 - 2.479X_5 - 5.281X_6 + 0.082\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.25)} = \begin{cases} 0.899 + 0.568X_{21} - 4.308X_{22} - 2.248X_3 - 0.324X_4 - 14.569X_5 - 2.756X_6 + 2.723\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.50)} = \begin{cases} -1.343 - 1.255X_{21} + 2.552X_{22} - 0.110X_4 - 0.356X_5 - 0.106X_6 + 2.986\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} -0.634 - 0.520X_{21} + 0.935X_{22} + 0.616X_4 - 0.065X_5 - 0.081X_6 + 2.473\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} -0.657 + 3.298X_{21} + 4.753X_{22} + 3.325X_4 - 0.453X_5 - 0.448X_6 + 11.116\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

Tabel 4.12 Bayes Faktor (2 Log B) Model Regresi TKB Reduksi Perkotaan

| Model Reduksi ( $M_2$ )                 | Kuantil              |                       |                        |                        |                         |
|---|----------------------|-----------------------|------------------------|------------------------|-------------------------|
|   | 0.05                 | 0.25                  | 0.50                   | 0.75                   | 0.95                    |
| Pendapatan rumah tangga ( $X_1$ )       | 0.005 <sup>(0)</sup> | 37.74 <sup>(**)</sup> | 434.52 <sup>(**)</sup> | 885.07 <sup>(**)</sup> | 1113.55 <sup>(**)</sup> |
| Pend. Menengah-rendah ( $X_{21}$ )      | 0.005 <sup>(0)</sup> | 37.74 <sup>(**)</sup> | 106.53 <sup>(**)</sup> | 604.53 <sup>(**)</sup> | 644.52 <sup>(**)</sup>  |
| Pend. Tinggi-rendah ( $X_{22}$ )        | 0.005 <sup>(0)</sup> | 24.9 <sup>(**)</sup>  | 120.08 <sup>(**)</sup> | 188.56 <sup>(**)</sup> | 186.76 <sup>(**)</sup>  |
| Persentase Pengeluaran pangan ( $X_3$ ) | 0.005 <sup>(0)</sup> | 9.44 <sup>(*)</sup>   | -44.03 <sup>(-)</sup>  | 28.72 <sup>(**)</sup>  | 32.36 <sup>(**)</sup>   |
| Jumlah anggota rumah tangga ( $X_4$ )   | 0.005 <sup>(0)</sup> | 20.49 <sup>(**)</sup> | -22.08 <sup>(-)</sup>  | -36.85 <sup>(-)</sup>  | -56.01 <sup>(-)</sup>   |
| Persentase ART bekerja ( $X_5$ )        | 0.005 <sup>(0)</sup> | 28.49 <sup>(**)</sup> | -56.52 <sup>(-)</sup>  | 86.29 <sup>(**)</sup>  | 45.21 <sup>(**)</sup>   |
| Persentase ART usia <12 tahun ( $X_6$ ) | 0.005 <sup>(0)</sup> | 37.74 <sup>(**)</sup> | 27.72 <sup>(**)</sup>  | 165.65 <sup>(**)</sup> | 86.57 <sup>(**)</sup>   |

Keterangan: Kekuatan pembuktian terhadap  $M_1$ , (-) : negatif, (0) : tidak ada; (+) : positif, (\*) : kuat, (\*\*): sangat kuat



Berdasarkan Tabel 4.12, besarnya Bayes faktor yang ada bagi masing-masing model regresi TKB menunjukkan bahwa model terbaik diperoleh apabila prediktor persentase pengeluaran pangan, jumlah anggota rumah tangga, dan persentase ART bekerja tidak dilibatkan dalam model kuantil 0.50. Adapun untuk model kuantil 0.75 dan 0.95 direkomendasikan agar meniadakan prediktor jumlah anggota rumah tangga. Di sisi lain, model kuantil 0.05 dan 0.25 sebaiknya menggunakan model penuh. Dari hasil ini, model kuantil 0.50 memiliki tiga kandidat model reduksi sebagai model terbaik, sehingga untuk analisis lanjutan dilakukan perhitungan bayes faktor lagi untuk model kuantil 0.50 tanpa mengikutsertakan prediktor persentase pengeluaran pangan, jumlah anggota rumah tangga, dan persentase ART bekerja.

Tabel 4.13 Bayes Faktor Model Regresi TKB Perkotaan Kuantil 0.50

| Model Reduksi ( $M_2$ )                 | Kuantil               |
|---|-----------------------|
|   | 0.50                  |
| Persentase Pengeluaran pangan ( $X_3$ ) | -76.85 <sup>(-)</sup> |
| Jumlah anggota rumah tangga ( $X_4$ )   |                       |
| Persentase ART bekerja ( $X_5$ )        |                       |

Keterangan: Kekuatan pembuktian terhadap  $M_1$ , (-) : negatif, (0) : tidak ada; (+) : positif; (\*) : kuat; (\*\*): sangat kuat

Tabel 4.13 memperlihatkan bahwa reduksi prediktor persentase pengeluaran pangan, jumlah anggota rumah tangga, dan persentase ART bekerja dari model regresi TKB kuantil 0.50 perkotaan menghasilkan model yang lebih baik daripada model penuh dan tiga model reduksi parsial sebelumnya. Oleh karena itu, dapat dikatakan ketiga prediktor tersebut tidak berperan dalam pembentukan model regresi TKB bagi rumah tangga di perkotaan yang memiliki pengeluaran untuk konsumsi susu sedang/rata-rata.

Di sisi lain, tidak seperti model wilayah pedesaan, nilai Bayes faktor model TKB wilayah perkotaan tidak mampu memberikan hasil yang pasti hanya



untuk kuantil 0.05. Kondisi ini menunjukkan bahwa estimasi regresi TKB bersifat kurang sensitif atau lemah dalam model pada kuantil-kuantil bawah khususnya kuantil 0.05. Hal ini juga sejalan dengan penemuan pada hasil simulasi untuk mengetahui performa estimator TKB sebelumnya.

Hasil estimasi parameter ulang untuk model-model reduksi ini terdapat dalam Lampiran 10. Adapun model-model regresi TKB terbaik akhir untuk pengeluaran konsumsi susu di perkotaan disajikan dalam persamaan-persamaan berikut.

$$\hat{y}_{(0.05)} = \begin{cases} -4.229 - 1.989 \log(X_1) + 1.012X_{21} + 3.056X_{22} - 5.158X_3 + 2.668X_4 - 4.562X_5 - 0.148X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.25)} = \begin{cases} -0.043 + 2.929 \log(X_1) + 0.206X_{21} + 0.941X_{22} - 0.041X_3 + 0.481X_4 - 1.341X_5 - 0.068X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.50)} = \begin{cases} -2.473 + 1.501 \log(X_1) + 1.323X_{21} + 4.289X_{22} - 0.073X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} 0.422 + 5.597 \log(X_1) + 0.477X_{21} + 1.420X_{22} - 0.118X_3 - 0.198X_5 - 0.242X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} 2.656 + 16.887 \log(X_1) + 1.355X_{21} + 6.811X_{22} - 0.308X_3 - 0.611X_5 - 0.405X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$



## DAFTAR LAMPIRAN

| Judul Lampiran   | Halaman |
|--|---------|
| Lampiran 1. Prosedur Pemrograman Linier Regresi Kuantil.....                 | 53      |
| Lampiran 2. Grafik RMSE Hasil Simulasi Jumlah Sampel.....                    | 55      |
| Lampiran 3. R Code Uji Beda (Uji t dan Uji <i>chi-square</i> ).....          | 57      |
| Lampiran 4. R-Code Uji Linieritas .....                                      | 58      |
| Lampiran 5. Hasil Uji Linieritas Transformasi $X_1$ dan $X_7$ .....          | 59      |
| Lampiran 6. R-Code Estimasi Parameter Model Regresi TKB .....                | 61      |
| Lampiran 7. R-Code Pemilihan Model Terbaik (Log marginal likelihood).....    | 63      |
| Lampiran 8. Log Marginal Likelihood Model TKB Pedesaan .....                 | 64      |
| Lampiran 9. Log Marginal Likelihood Model TKB Perkotaan .....                | 66      |
| Lampiran 10. Hasil Estimasi Parameter Ulang Pemilihan Model TKB Terbaik..... | 68      |



## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut

1. Fungsi likelihood pada persamaan berikut digunakan untuk memperoleh estimator model regresi TKB dengan pendekatan bayesian (teknik MCMC Metropolis-Hastings).

$$L(\mathbf{y}|\boldsymbol{\beta}(\theta)) = \theta^n (1-\theta)^n \exp \left\{ -\rho_{\theta} \sum_{i=1}^n (y_i - \max(\mathbf{x}_i' \boldsymbol{\beta}, 0)) \right\}$$

Di sisi lain, pemilihan model terbaik regresi TKB menggunakan formulasi Bayes faktor berikut ini.

$$B_{12} = \exp \left\{ \log \hat{m}(\mathbf{y}|M_1) - \log \hat{m}(\mathbf{y}|M_2) \right\}$$

2. Berdasarkan hasil simulasi, performa estimator regresi TKB lebih baik daripada estimator Powell ketika ukuran sampel yang digunakan lebih dari 1000 dan prediktor yang digunakan cukup banyak. Di sisi lain, kedua estimator tersebut ternyata tidak cukup baik untuk memodelkan kuantil bawah seperti kuantil 0.05. Selain itu, performa estimator TKB dan Powell lebih baik daripada estimator tobit standar untuk model yang memiliki error berdistribusi bukan normal.
3. Model regresi TKB terbaik untuk pengeluaran konsumsi susu di wilayah pedesaan ditunjukkan oleh persamaan-persamaan berikut.

$$\hat{y}_{(0.05)} = \begin{cases} -0.464 - 2.633X_{21} + 1.288X_{22} - 4.103X_3 + 0.374X_4 - 2.479X_5 - 5.281X_6 + 0.082\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$
$$\hat{y}_{(0.25)} = \begin{cases} 0.899 + 0.568X_{21} - 4.308X_{22} - 2.248X_3 - 0.324X_4 - 14.569X_5 - 2.756X_6 + 2.723\log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$



$$\hat{y}_{(0.50)} = \begin{cases} -1.343 - 1.255X_{21} + 2.552X_{22} - 0.110X_4 - 0.356X_5 - 0.106X_6 + 2.986 \log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} -0.634 - 0.520X_{21} + 0.935X_{22} + 0.616X_4 - 0.065X_5 - 0.081X_6 + 2.473 \log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} -0.657 + 3.298X_{21} + 4.753X_{22} + 3.325X_4 - 0.453X_5 - 0.448X_6 + 11.116 \log(X_7) & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

Adapun model regresi TKB terbaik pengeluaran rumah tangga untuk konsumsi susu di wilayah perkotaan sebagai berikut.

$$\hat{y}_{(0.05)} = \begin{cases} -4.229 - 1.989 \log(X_1) + 1.012X_{21} + 3.056X_{22} - 5.158X_3 + 2.668X_4 - 4.562X_5 - 0.148X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.25)} = \begin{cases} -0.043 + 2.929 \log(X_1) + 0.206X_{21} + 0.941X_{22} - 0.041X_3 + 0.481X_4 - 1.341X_5 - 0.068X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.50)} = \begin{cases} -2.473 + 1.501 \log(X_1) + 1.323X_{21} + 4.289X_{22} - 0.073X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.75)} = \begin{cases} 0.422 + 5.597 \log(X_1) + 0.477X_{21} + 1.420X_{22} - 0.118X_3 - 0.198X_5 - 0.242X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

$$\hat{y}_{(0.95)} = \begin{cases} 2.656 + 16.887 \log(X_1) + 1.355X_{21} + 6.811X_{22} - 0.308X_3 - 0.611X_5 - 0.405X_6 & y^* > 0 \\ 0 & y^* \leq 0 \end{cases}$$

## 5.2 Saran

Saran-saran yang dapat disampaikan berdasarkan hasil penelitian yang ada adalah sebagai berikut.

1. Pada penelitian ini, model tobit yang digunakan bersifat linier sehingga prediktor yang bersifat nonlinier namun memiliki peranan penting, seperti pendapatan rumah tangga, yang tidak dapat dimasukkan dalam model.





Oleh karena itu diperlukan pengembangan terhadap model regresi TKB nonlinier.

2. Dari hasil simulasi dan analisis yang telah dilakukan, ditemukan bahwa estimator regresi TKB tidak cukup baik untuk mengestimasi model untuk kuantil-kuantil bawah. Oleh karena itu disarankan untuk melakukan modifikasi dan kajian lebih lanjut terkait hal tersebut.



## DAFTAR PUSTAKA

- Agresti, A. (2002). *Categorical Data Analysis*. New York: John Wiley & Sons.
- Aisyah, N., Arumugam, N., Hussein, M. A., dan Latiff, I. (2012). "Factors Affecting the Technical Efficiency Level of Inshore Fisheries in Kuala Trengganu Malaysia". *International Journal of Agriculture Management and Development*, Vol. 1, 49-56.
- Al-Hassan, Y. (2010). "Performance of a New Ridge Regression Estimator ". *Journal of the Association of Arab Universities for Basic and Applied Sciences*, Vol. 9, 43-50.
- Alwis, A. E. N. D., Ediringhe, J. C., dan Athauda, A. M. T. P. (2009). "Analysis of Factors Affecting Fresh Milk Consumption Among The Mid-Country Consumers". *Tropical Agricultural Research Extension*, Vol. 12, 101-107.
- Babolian, H. R., dan Karim, M. S. A. (2010). "Factors Affecting Milk Consumption Among School Children In Urban and Rural Areas of Selangor Malaysia". *International Food Research Journal*, Vol. 17, 591-601.
- Biliyas, Y., Chen, S., dan Ying, Z. (2000). "Simple Resampling Methods for Censored Regression Quantiles". *Journal of Econometrics*, Vol. 68, 303-338.
- Buchinsky, M. (1994). "Changes in U.S. Wage Structure 1963-1987: Applications of Quantile Regression". *Econometrica*, Vol. 62, 405-458.
- Buhai, I. S. (2005). "Quantile Regression: Overview and Selected Applications". *Ad Astra*, Vol. 4, 1-17.
- Chai, T., dan Draxler, R. R. (2014). "Root Mean Square Error (RMSE) or Mean Absolute Error (MAE): Arguments Against Avoiding RMSE in The Literature". *Geoscientific Model Development*, Vol. 7, 1247-1250.
- Chen, C. (2005). An Introduction to Quantile Regression and The QUANTREG Procedure. Retrieved October 20<sup>th</sup>, 2014, from <http://www2.sas.com/proceeding/sugi30/213-30.pdf>
- Chib, S., dan Jeliakoz, I. (2001). "Marginal Likelihood from Metropolis-Hastings Algorithm". *Journal of The American Statistical Association*, Vol. 96, 270-281.
- Fitzenberger, B. (1997). *Computational Aspects of Censored Quantile Regression*. Paper presented at the 3rd International Data Analysis based on L- 1- Statistical Procedures and Related Methods, Hayward, California.
- Fitzenberger, B., dan Winker, P. (2007). "Improving the Computation of Censored Quantile Regression Estimators". *CSDA*, Vol. 52, 88-108.



- Gelman, A., Carlin, J. B., Stern, H. S., dan Rubin, D. B. (2004). *Bayesian Data Analysis Second Edition*. Florida: Chapman and Hall.
- Gheblawi, M., dan Sherif, S. (2007). "Determination of Factors Affecting Expenditures on Three Major Food Groups in Al-Ain The United Arab Emirates (UAE)". *Emirates Journal of Food and Agriculture*, Vol. 19, 15-23.
- Glass, G., dan Hopkins, K. (1995). *Statistical Methods in Education and Psychology 3rd Edition*. Boston: Allyn and Bacon.
- Greene, W. H. (2001). *Econometrics Analysis 5th Edition*. New Jersey: Prentice Hall.
- Kass, R. E., dan Raftery, A. E. (1995). "Bayes Factors". *Journal of The American Statistical Association*, Vol. 90, 773-795.
- Koenker, R., dan Basset, G. (1978). "Regression Quantiles". *Econometrica*, Vol. 46, 33-50.
- Koenker, R., dan Hallock, K. (2001). "Quantile Regressions". *Journal of Economic Perspectives*, Vol. 15, 143-156.
- Koenker, R., dan Machado, J. A. F. (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression". *Journal of The American Statistical Association*, Vol. 94, 1296-1310.
- Komrattanapanya, P. (2013). "Factors Influencing Dividen Payment in Thailand: A Tobit Regression Analysis". *International Journal of Accounting and Financial Reporting*, Vol. 3, 225-268.
- Lavine, M., dan Schervish, M. J. (1999). "Bayes Factors: What They Are and What They Are Not". *The American Statistician*, Vol. 53, 119-122.
- Lehmann, E. L., dan Casella, G. (1999). *Theory of Point Estimation*. New York: Springer-Verlag Inc.
- Leiker, A. (2012). *A Comparison Study On The Estimation In Tobit Regression Model* (D. o. S. C. A. a. Science, Trans.). Kansas: Kansas State University.
- McBee, M. (2010). "Modelling Outcomes With Floor or Ceiling Effects: An Introduction to Tobit Model". *Gifted Child Quarterly*, Vol. 54, 313-320.
- Osborne, J., dan Waters, E. (2002). "Four Assumptions Of Multiple Regression That Researchers Should Always Test". *Practical Assessment, Research & Evaluation*, Vol. 8 (1), 1-5.
- Phuong, N. V., Mergenthaler, M., dan Tran, C. H. (2013). *Effects of Socio-economic and Demography Variables on Vietnamese Household' Expenditure for Dairy Product*. Paper presented at the The Tropentag



Symposium: Agricultural Development Within The Rural-Urban Continuum Stuttgart.

Powell, J. (1986). "Censored Regression Quantiles". *Journal of Econometrics*, Vol. 32, 143-155.

Rahardja, P., dan Manurung, M. (2008). *Pengantar Ilmu Ekonomi (Mikroekonomi & Makroekonomi) Edisi Ketiga*. Jakarta: FE UI.

Robinson, C., Tomek, S., dan Schumaker, R. (2013). "Test of Moderation Effects: Difference in Simple Slopes versus the Interaction Term". *Multiple Linear Regression Viewpoints*, Vol. 39 (16-24).

Sekhampu, T. J. (2012). "Socio-economic Determinants of Household Food Expenditure in A Low Income Township in South Africa". *Mediterranean Journal of Social Sciences*, Vol. 3, 449-453.

Shukur, G., dan Mantalos, P. (2004). "Size and Power of the RESET Test as Applied to Systems of Equations: A Bootstrap Approach". *Journal of Modern Applied Statistical Methods*, Vol. 3 (2), 1-22.

Sukirno, S. (2004). *Teori dan Pengantar Makro Ekonomi*. Jakarta: Grafindo Persada.

Tobin, J. (1958). "Estimation of Relationship for Limited Dependent Variables". *Econometrica*, Vol. 26, 24-36.

Trung, T. Q., Giam, D. Q., Hai, V. T., Thao, L. P., Hang, N. T. T., Son, L. T. K., dan Linh, B. T. M. (2014). "Factors Influencing Milk Consumption of Rural Households in Northern Vietnam". *Greener Journal of Business and Management Studies*, Vol. 4, 31-40.

USDA. (2013). Indonesian Dairy and Product Annual 2013. from [http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Dairy%20and%20Products%20Annual\\_Jakarta\\_Indonesia\\_11-22-2013.pdf](http://gain.fas.usda.gov/Recent%20GAIN%20Publications/Dairy%20and%20Products%20Annual_Jakarta_Indonesia_11-22-2013.pdf)

Uzunoz, M., dan Akcay, Y. (2012). "A Case Study of Probit Model Analysis of Factors Affecting Consumption of Packed and Unpacked Milk in Turkey". *Economic Research International Journal*, Vol. 1, 1-8.

Wang, M., dan Zhang, L. (2012). "A Bayesian Quantile Regression Analysis of Potential Risk Factors for Violent Crimes in USA". *Open Journal of Statistics*, Vol. 2, 73-78.

Wham, C. A., dan Worsley, A. (2003). "New Zealanders' Attitudes To Milk: Implication for Public Health". *Public Health Nutrition*, Vol. 6, 73-78.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Massachusetts: MIT Press.



Yao, Y., dan Lee, Y. (2010). Another Look at Linear Programming for Feature Selection via Methods of Regularization. Ohio: The Ohio State University.

Yu, K., dan Moyeed, R. A. (2001). "Bayesian Quantile Regression". *Statistics and Probability Letters*, Vol. 54, 437-447.

Yu, K., dan Stander, J. (2007). "Bayesian Analysis of Tobit Quantile Regression Model". *Journal of Econometrics*, Vol. 137, 260-276.

Yue, Y. R., dan Hong, H. G. (2012). "Bayesian Tobit Quantile Regression Model for Medical Expenditure Panel Survey Data". *Statistical Modelling*, Vol. 12, 323-346.

Zain, I. (1997). *Model Regresi Tobit dan Aplikasinya*. Surabaya: Lembaga Penelitian Institut Teknologi Sepuluh Nopember.

Zain, I., Widodo, D. A., dan Ratnasari, V. (2011). Pengembangan Model Data Tersensor Dengan Metode Regresi Tobit Spasial Bivariat Surabaya: Institut Teknologi Sepuluh Nopember.



## LAMPIRAN - LAMPIRAN

### Lampiran 1. Prosedur Pemrograman Linier Regresi Kuantil

Berikut ini merupakan formulasi pemrograman linier untuk model regresi kuantil dimana bentuk persamaan (2.17) mengalami sedikit modifikasi sebagaimana ditunjukkan persamaan berikut (Yao dan Lee, 2010)

$$\begin{aligned} \min_{\beta \in \mathbb{R}^p} & \left[ \frac{1}{n} \sum_{i=1}^n \rho_{\theta}(y_i - f(\mathbf{x}_i)) \right] \\ \min_{\beta \in \mathbb{R}^p} & \left[ \left( \theta - \frac{1}{2} \right) (\bar{y} - \bar{\mathbf{x}}' \boldsymbol{\beta}) + \frac{1}{2n} \sum_{i=1}^n |y_i - \mathbf{x}_i' \boldsymbol{\beta}| \right] \\ \max & \sum_{i=1}^n y_i a_i \quad \text{s.t. } \mathbf{x}' a = (1 - \theta) \mathbf{x}' \mathbf{1} \quad \text{dan } a \in [0, 1]^n \end{aligned}$$

di mana:  $\rho_{\theta}(u) = \theta u^+ + (1 - \theta) u^-$

Bentuk umum pemrograman linier yaitu

$$\begin{aligned} \min & \mathbf{c}' \mathbf{z} \\ \text{s.t.} & \mathbf{A} \mathbf{z} = \mathbf{b} \quad \text{dan } \mathbf{z} \geq \mathbf{0} \end{aligned}$$

Model regresi kuantil sebagai permasalahan pemrograman linier yaitu

$$\mathbf{z} = (\boldsymbol{\beta}^+ \quad \boldsymbol{\beta}^- \quad \boldsymbol{\varepsilon}^+ \quad \boldsymbol{\varepsilon}^-)^t$$

$$\mathbf{c} = \left( 0 \quad 0 \quad \frac{\theta}{n} \quad \frac{1-\theta}{n} \right)^t$$

$$\mathbf{A} = (\mathbf{x} \quad -\mathbf{x} \quad \mathbf{I} \quad -\mathbf{I})$$

Dengan  $\mathbf{c}$  dan  $\mathbf{z}$  adalah vektor berukuran  $m=2p+2n$ ,  $\mathbf{A}$  adalah matriks berukuran  $n \times m$  dan  $\boldsymbol{\varepsilon}$  adalah vektor residual.

Proses iterasi metode simpleks adalah sebagai berikut.



1. Pandang  $\mathbf{B} \equiv \{\mathbf{B}_1, \dots, \mathbf{B}_n\} \subset \{1, 2, \dots, n\}$  sebagai himpunan indeks basis,

$\mathbf{A}_B \equiv [\mathbf{A}_{B_1}, \dots, \mathbf{A}_{B_n}]$  sebagai sub-matriks  $\mathbf{A}$  yang bersifat full-rank, dan

$\mathbf{z}^* = (z_1^*, \dots, z_m^*)$  disebut sebagai penyelesaian awal jika  $\mathbf{z}^*$  memenuhi kriteria

$$\mathbf{z}_B^* = \mathbf{A}_B^{-1} \mathbf{b}$$

$$\mathbf{z}_j^* = \mathbf{0} \text{ untuk } j \in \{1, 2, \dots, m\} \setminus B$$

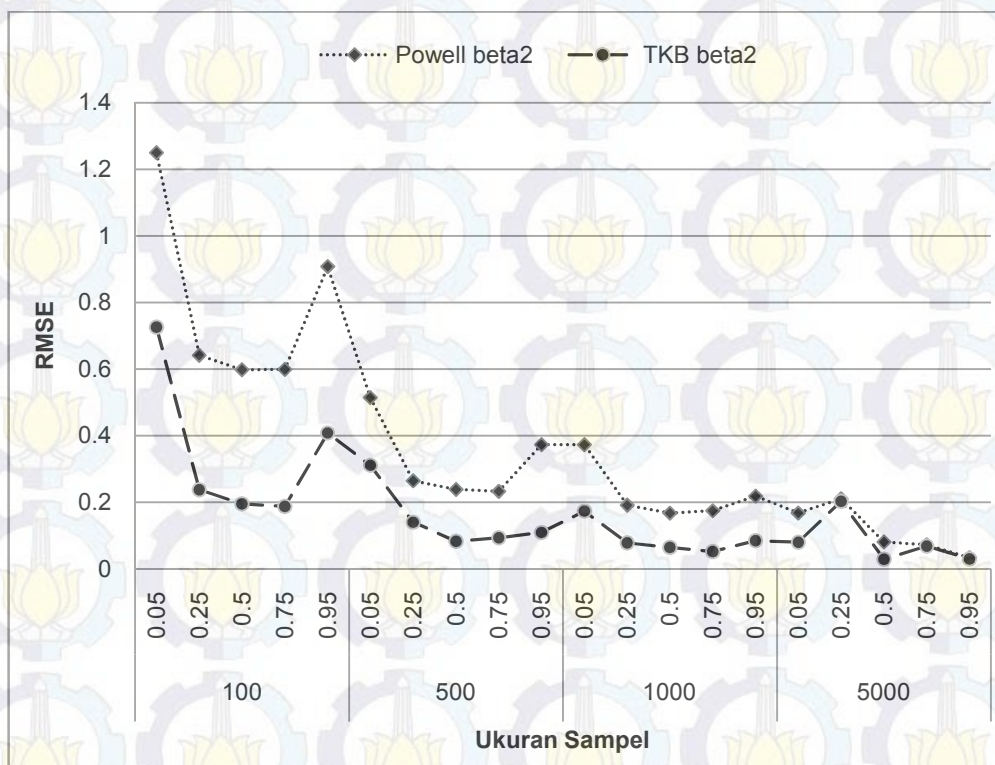
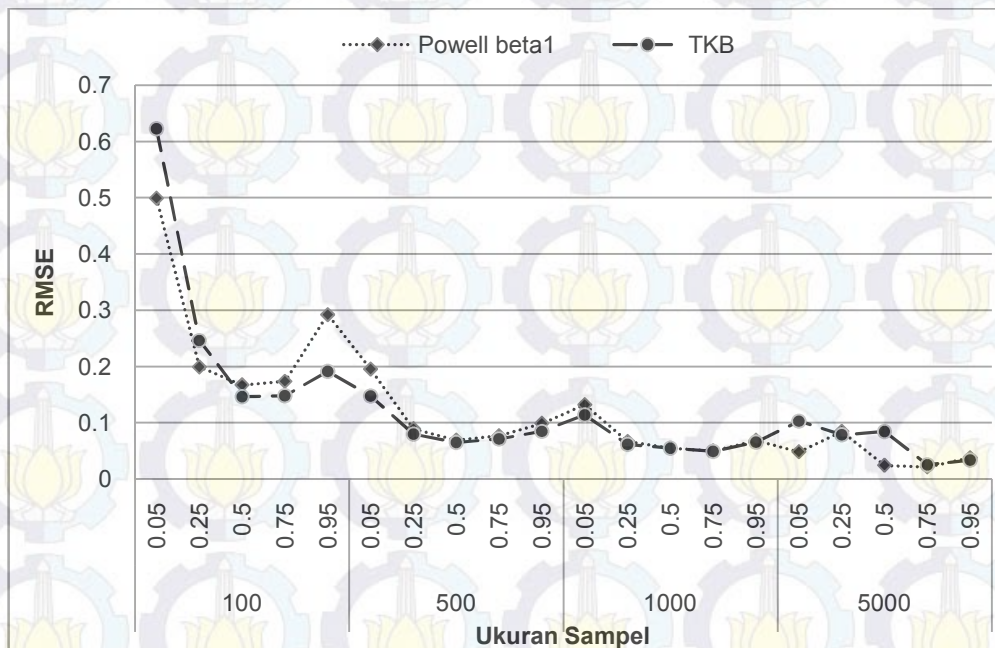
2.  $\mathbf{z}^*$  memenuhi kriteria adalah penyelesaian optimal apabila  $\mathbf{z}_B^* = \mathbf{A}_B^{-1} \mathbf{b} \geq \mathbf{0}$  dan  $\mathbf{c} - \mathbf{A}' \mathbf{A}_B^{-1} \mathbf{c}_B \geq \mathbf{0}$  maka  $\mathbf{z}^*$  adalah solusi optimal. Namun jika tidak terpenuhi maka  $\mathbf{z}^*$  proses akan menjadi penyelesaian awal lagi dan kembali ke langkah 1 sampai kondisi optimal tercapai.

3. Tabel simpleks

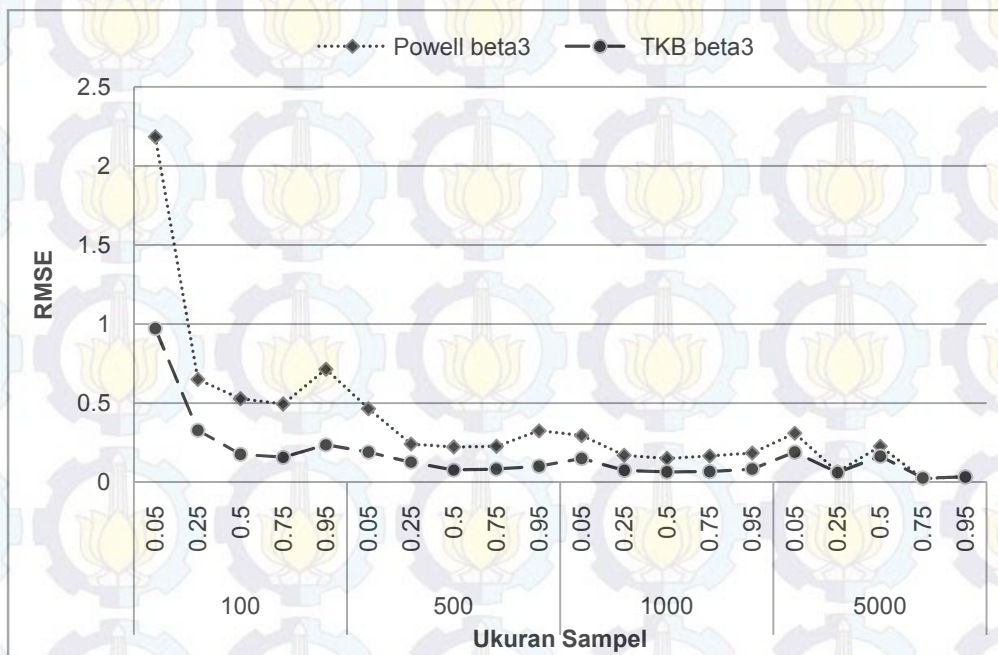
$$\begin{bmatrix} -\mathbf{c}_B' \mathbf{z}_B & \mathbf{c}' - \mathbf{c}_B' \mathbf{A}_B^{-1} \mathbf{A} \\ \mathbf{A}_B^{-1} \mathbf{b} & \mathbf{A}_B^{-1} \mathbf{A} \end{bmatrix}$$



**Lampiran 2. Grafik RMSE Hasil Simulasi Jumlah Sampel**









### Lampiran 3. R Code Uji Beda (Uji t dan Uji *chi-square*)

```
data=read.table("d://data fix.txt",header=TRUE)
x1=as.numeric(data$X1)
x21=as.numeric(data$d21)
x22=as.numeric(data$d22)
x3=as.numeric(data$X3)
x4=as.numeric(data$X4)
x5=as.numeric(data$X5)
x6=as.numeric(data$X6)
x7=as.numeric(data$X7)
y=as.numeric(data$Y)
wil=as.numeric(data$wil)
t.test(x1~wil)
t.test(x3~wil)
t.test(x4~wil)
t.test(x5~wil)
t.test(x6~wil)
t.test(x7~wil)
t.test(y~wil)
library(MASS)
tab1=table(x21,wil)
tab2=table(x22,wil)
chisq.test(tab1)
chisq.test(tab2)
```



#### Lampiran 4. R-Code Uji Linieritas

```
data=read.table("d://data desa kon r.txt",header=TRUE)
x1=as.numeric(data$X1)
d21=as.numeric(data$d21)
d22=as.numeric(data$d22)
x3=as.numeric(data$X3)
x4=as.numeric(data$X4)
x5=as.numeric(data$X5)
x6=as.numeric(data$X6)
x7=as.numeric(data$X7)

y=as.numeric(data$Y)
library(lmtest)
out1=reset(y~x1, power=3)
out3=reset(y~x3, power=3)
out4=reset(y~x4, power=3)
out5=reset(y~x5, power=3)
out6=reset(y~x6, power=3)
out7=reset(y~x7, power=3)
out21=lm(y~d21)
out22=lm(y~d22)
out1
out3
out4
out5
out6
out7
summary(out21)
summary(out22)
```



## Lampiran 5. Hasil Uji Linieritas Transformasi $X_1$ dan $X_7$

### - Wilayah Pedesaan

```
> data=read.table("d://data desa kon r.txt",header=TRUE)
> x1=as.numeric(data$X1)
> d21=as.numeric(data$d21)
> d22=as.numeric(data$d22)
> x3=as.numeric(data$X3)
> x4=as.numeric(data$X4)
> x5=as.numeric(data$X5)
> x6=as.numeric(data$X6)
> x7=as.numeric(data$X7)
> y=as.numeric(data$Y)
> library(lmtest)
> out11=reset(y~log(x1), power=3)
> out12=reset(y~(x1^2), power=3)
> x1=x1^0.5
> out13=reset(y~x1, power=3)
> out71=reset(y~log(x7), power=3)
> out72=reset(y~(x7^2), power=3)
> x7=x7^0.5
> out73=reset(y~x7, power=3)
> out11
```

RESET test

```
data: y ~ log(x1)
RESET = 182.8184, df1 = 1, df2 = 1570, p-value < 2.2e-16
```

```
> out12
```

RESET test

```
data: y ~ (x1^2)
RESET = 7.0772, df1 = 1, df2 = 1570, p-value = 0.007887
```

```
> out13
```

RESET test

```
data: y ~ x1
RESET = 76.0906, df1 = 1, df2 = 1570, p-value < 2.2e-16
```

```
> out71
```

RESET test

```
data: y ~ log(x7)
RESET = 0.8394, df1 = 1, df2 = 1570, p-value = 0.3597
```

```
> out72
```

RESET test

```
data: y ~ (x7^2)
RESET = 1.7371, df1 = 1, df2 = 1570, p-value = 0.1877
```

```
> out73
```

RESET test

```
data: y ~ x7
RESET = 0.1707, df1 = 1, df2 = 1570, p-value = 0.6796
```

### - Wilayah Perkotaan



```

> data=read.table("d://data kota kon r.txt",header=TRUE)
> x1=as.numeric(data$X1)
> d21=as.numeric(data$d21)
> d22=as.numeric(data$d22)
> x3=as.numeric(data$X3)
> x4=as.numeric(data$X4)
> x5=as.numeric(data$X5)
> x6=as.numeric(data$X6)
> x7=as.numeric(data$X7)
> y=as.numeric(data$Y)
> library(lmtest)
> out11=reset(y~log(x1), power=3)
> out12=reset(y~(x1^2), power=3)
> x1=x1^0.5
> out13=reset(y~x1, power=3)
> out71=reset(y~log(x7), power=3)
> out72=reset(y~(x7^2), power=3)
> x7=x7^0.5
> out73=reset(y~x7, power=3)
> out11

```

RESET test

```

data: y ~ log(x1)
RESET = 2.0925, df1 = 1, df2 = 1240, p-value = 0.1483
> out12

```

RESET test

```

data: y ~ (x1^2)
RESET = 48.1312, df1 = 1, df2 = 1240, p-value = 6.405e-12
> out13

```

RESET test

```

data: y ~ x1
RESET = 2.0632, df1 = 1, df2 = 1240, p-value = 0.1511
> out71

```

RESET test

```

data: y ~ log(x7)
RESET = 11.0326, df1 = 1, df2 = 1240, p-value = 0.0009213
> out72

```

RESET test

```

data: y ~ (x7^2)
RESET = 24.3966, df1 = 1, df2 = 1240, p-value = 8.912e-07
> out73

```

RESET test

```

data: y ~ x7
RESET = 37.0661, df1 = 1, df2 = 1240, p-value = 1.52e-09

```



## Lampiran 6. R-Code Estimasi Parameter Model Regresi TKB

```
library(lqmm)
library(coda)
data=read.table("d://data fix.txt",header=TRUE)
x1=as.numeric(data$X1)
x21=as.numeric(data$d21)
x22=as.numeric(data$d22)
x3=as.numeric(data$X3)
x4=as.numeric(data$X4)
x5=as.numeric(data$X5)
x6=as.numeric(data$X6)
x7=as.numeric(data$X7)
y=as.numeric(data$Y)

## mendefinisikan fungsi distribusi asimetris Laplace
ald=function(a,b)
{
  u=a-b
  ll=dal(u, 0, 1, 0.95, log=T)
}

## model regresi tobit
likelihood<-function(param)
{
  a=param[1]
  b1=param[2]
  b2=param[3]
  b3=param[4]
  b4=param[5]
  b5=param[6]
  b6=param[7]
  b7=param[8]
  pred1 <- a+b1*x21+b2*x22+x3*b3+x4*b4+x5*b5+x6*b6+x8*b7
  pred=pmax(pred1,0)
  sumll<-sum(ald(y,pred))
  return(sumll)
}

## mendefinisikan distribusi prior
prior<-function(param)
{
  a=param[1]
  b1=param[2]
  b2=param[3]
  b3=param[4]
  b4=param[5]
  b5=param[6]
  b6=param[7]
  b7=param[8]
  aprior<-dunif(a,-100,100) b1prior<-dunif(b1,-100,100)
  b2prior<-dunif(b2,-100,100)
  b3prior<-dunif(b3,-100,100)
  b4prior<-dunif(b4,-100,100)
  b5prior<-dunif(b5,-100,100)
```



```

b6prior<-dunif(b6,-100,100)
b7prior<-dunif(b7,-100,100)
return(aprior+b1prior+b2prior+b3prior+b4prior+b5prior+b6prior+7prior)
}

proposalfunction <- function(param){
  return(rnorm(8,mean = param, sd= c(1,1,1,1,1,1,1,1)))
}

run_metropolis_MCMC <- function(startvalue, iterations){
  chain = array(dim = c(iterations+1,8))
  lik= array(dim = c(iterations+1,1))
  chain[1,] = startvalue
  for (i in 1:iterations){
    proposal = proposalfunction(chain[i,])
    probab = exp(likelihood(proposal)+ prior(proposal) -
      likelihood(chain[i,]) - prior(chain[i,]))
    if (runif(1) < probab){
      chain[i+1,] = proposal
    }else{
      chain[i+1,] = chain[i,]
      lik[i,]=likelihood(chain[i,])
    }
  }
  return(mcmc(chain))
}

startvalue=c(0,0,0,0,0,0,0,0)
chain=run_metropolis_MCMC(startvalue, 5000)
summary(chain)
burnIn=1000
qtb=mean(chain[-(1:burnIn),])

```



## Lampiran 7. R-Code Pemilihan Model Terbaik (Log marginal likelihood)

```
like=function(chain)
{
  k=nrow(chain)
  lik=data.frame()
  for (i in 1:k)
  {
    lik.temp=likelihood(chain[i,])
    lik=rbind(lik, lik.temp)
  }
  result=lik
}
lik=like(chain)
lik=as.matrix(lik)
summary(lik)

marginal.likelihood <- function(x,lik, num.samples=1000,log=TRUE) {
  if (class(x) != "mcmc" & class(x) != "mcmc.list") {
    stop("x harus merupakan objek mcmc atau mcmcList objek.")
  }
  y1 <- summary(x)
  param.star <- y1$statistics[, "Mean"]
  lik.star <- likelihood(param.star)
  g <- sample(1:nrow(x), num.samples, replace=TRUE)
  param.g <- x[g,]
  lik.g <- lik[g]
  alpha.g <-
    sapply(lik.g, function(l) min(1, prior(param.star)/prior(param.g)*(lik.
      star/lik.g)))
  param.j <- proposalfunction(param.star)
  lik.j <- likelihood(param.j)
  alpha.j <- sapply(lik.j, function(l)
    min(1, prior(param.j)/prior(param.star)*(lik.j/lik.star)))
  pi.hat <- mean(alpha.g)/mean(alpha.j)
  pi.star <- prior(param.star)
  ln.m <- lik.star + pi.star - log(pi.hat)
  list(ln.m=ln.m, aj=alpha.j, ag=alpha.g[1],
    ln.lik.star=lik.star, ln.pi.star=pi.star, ln.pi.hat=log(pi.hat)
  )
}
x=chain
lik=lik
m=marginal.likelihood(x,lik)
```



## Lampiran 8. Log Marginal Likelihood Model TKB Pedesaan

- Model Penuh

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.24           |
| 0.25 | -13505.61           |
| 0.50 | -19960.86           |
| 0.75 | -24148              |
| 0.95 | -21909.71           |

- Model reduksi  $X_{21}$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -20097.25           |
| 0.75 | -24901.67           |
| 0.95 | -22731.31           |

- Model reduksi  $X_{22}$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -19975.15           |
| 0.75 | -24340.49           |
| 0.95 | -21958.92           |

- Model reduksi  $X_3$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -19889.02           |
| 0.75 | -24121.67           |
| 0.95 | -21729.47           |

- Model reduksi  $X_4$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -20183.18           |
| 0.75 | -24455.48           |
| 0.95 | -22431.57           |



- Model reduksi  $X_5$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -20051.28           |
| 0.75 | -24200.03           |
| 0.95 | -22054.03           |

- Model reduksi  $X_6$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -19989.33           |
| 0.75 | -24379.29           |
| 0.95 | -22779.68           |

- Model reduksi  $X_7$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -12115.25           |
| 0.25 | -13505.62           |
| 0.50 | -20170.13           |
| 0.75 | -24978.28           |
| 0.95 | -22952.37           |



## Lampiran 9. Log Marginal Likelihood Model TKB Perkotaan

- Model Penuh

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.563           |
| 0.25 | -12726.82           |
| 0.50 | -20146.3            |
| 0.75 | -23514.02           |
| 0.95 | -17229.79           |

- Model reduksi  $X_1$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12764.56           |
| 0.50 | -20580.82           |
| 0.75 | -24399.09           |
| 0.95 | -18343.34           |

- Model reduksi  $X_{21}$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12764.56           |
| 0.50 | -20252.83           |
| 0.75 | -24118.55           |
| 0.95 | -17874.31           |

- Model reduksi  $X_{22}$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12751.72           |
| 0.50 | -20266.38           |
| 0.75 | -23702.58           |
| 0.95 | -17416.55           |

- Model reduksi  $X_3$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12736.26           |
| 0.50 | -20102.27           |
| 0.75 | -23542.74           |
| 0.95 | -17262.15           |



- Model reduksi  $X_4$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12747.31           |
| 0.50 | -20124.22           |
| 0.75 | -23477.17           |
| 0.95 | -17173.78           |

- Model reduksi  $X_5$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12755.31           |
| 0.50 | -20089.78           |
| 0.75 | -23600.31           |
| 0.95 | -17275              |

- Model reduksi  $X_6$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.05 | -7871.568           |
| 0.25 | -12764.56           |
| 0.50 | -20174.02           |
| 0.75 | -23679.67           |
| 0.95 | -17316.36           |

- Model reduksi  $X_3, X_4, X_5$

| Q    | Marginal likelihood |
|------|---------------------|
| 0.50 | -20069.45           |



## Lampiran 10. Hasil Estimasi Parameter Ulang Pemilihan Model TKB Terbaik

- Model regresi TKB kuantil 0.50 pedesaan

|      | Mean    | SD      | Naive SE  | Time-series SE |
|------|---------|---------|-----------|----------------|
| [1,] | -1.3435 | 0.67608 | 0.0095603 | 0.36624        |
| [2,] | -1.2547 | 0.19353 | 0.0027367 | 0.04060        |
| [3,] | 2.5522  | 1.53013 | 0.0216372 | 1.21276        |
| [4,] | -0.1100 | 0.17024 | 0.0024074 | 0.05163        |
| [5,] | -0.3555 | 0.06055 | 0.0008562 | 0.02263        |
| [6,] | -0.1057 | 0.07242 | 0.0010240 | 0.01075        |
| [7,] | 2.9861  | 0.43939 | 0.0062132 | 0.12925        |

- Model regresi TKB kuantil 0.75 pedesaan

|      | Mean     | SD      | Naive SE  | Time-series SE |
|------|----------|---------|-----------|----------------|
| [1,] | -0.63371 | 0.19518 | 0.0027599 | 0.08184        |
| [2,] | -0.52035 | 0.22735 | 0.0032149 | 0.14125        |
| [3,] | 0.93538  | 0.40932 | 0.0057880 | 0.25926        |
| [4,] | 0.61577  | 0.25784 | 0.0036461 | 0.12001        |
| [5,] | -0.06549 | 0.04615 | 0.0006527 | 0.01597        |
| [6,] | -0.08120 | 0.04890 | 0.0006914 | 0.01749        |
| [7,] | 2.47359  | 0.73928 | 0.0104540 | 0.43719        |

- Model regresi TKB kuantil 0.95 pedesaan

|      | Mean    | SD     | Naive SE | Time-series SE |
|------|---------|--------|----------|----------------|
| [1,] | -0.6570 | 1.4030 | 0.019839 | 1.41361        |
| [2,] | 3.2980  | 1.7145 | 0.024244 | 1.30278        |
| [3,] | 4.7532  | 1.2457 | 0.017615 | 0.57841        |
| [4,] | 3.3246  | 0.7702 | 0.010891 | 0.34366        |
| [5,] | -0.4532 | 0.2295 | 0.003245 | 0.09762        |
| [6,] | -0.4479 | 0.1677 | 0.002371 | 0.06609        |
| [7,] | 11.1164 | 3.4010 | 0.048093 | 1.98608        |

- Model regresi TKB kuantil 0.50 perkotaan

|      | Mean     | SD      | Naive SE | Time-series SE |
|------|----------|---------|----------|----------------|
| [1,] | -2.47254 | 2.01233 | 0.028456 | 1.457075       |
| [2,] | 1.50071  | 0.30335 | 0.004290 | 0.086293       |
| [3,] | 1.32278  | 0.76187 | 0.010773 | 0.507819       |
| [4,] | 4.28879  | 2.58044 | 0.036489 | 2.075383       |
| [5,] | -0.07262 | 0.03628 | 0.000513 | 0.007834       |



- Model regresi TKB kuantil 0.75 perkotaan

|      | Mean    | SD Naive | SE Time-series | SE      |
|------|---------|----------|----------------|---------|
| [1,] | 0.4221  | 0.21888  | 0.003095       | 0.08433 |
| [2,] | 5.5971  | 1.81695  | 0.025693       | 1.09944 |
| [3,] | 0.4767  | 0.30625  | 0.004331       | 0.13357 |
| [4,] | 1.4196  | 0.92365  | 0.013061       | 0.77588 |
| [5,] | -0.1176 | 0.08896  | 0.001258       | 0.03324 |
| [6,] | -0.1976 | 0.14466  | 0.002046       | 0.07912 |
| [7,] | -0.2418 | 0.10583  | 0.001497       | 0.04519 |

- Model regresi TKB kuantil 0.95 perkotaan

|      | Mean    | SD Naive | SE Time-series | SE      |
|------|---------|----------|----------------|---------|
| [1,] | 2.6558  | 0.8707   | 0.012312       | 0.48429 |
| [2,] | 16.8871 | 4.9173   | 0.069534       | 2.78894 |
| [3,] | 1.3553  | 1.0115   | 0.014303       | 0.79493 |
| [4,] | 6.8109  | 2.7003   | 0.038184       | 1.90248 |
| [5,] | -0.3077 | 0.2201   | 0.003112       | 0.07068 |
| [6,] | -0.6113 | 0.5077   | 0.007180       | 0.31518 |
| [7,] | -0.4052 | 0.2940   | 0.004157       | 0.17912 |



## BIOGRAFI PENULIS



Penulis lahir di Kota Kupang, Provinsi Nusa Tenggara Timur pada tanggal 24 Juni 1993 dengan nama lengkap Evellin Dewi Lusiana, sebagai anak pertama dari dua bersaudara dari pasangan Sunadji dan Sumianti. Penulis menempuh pendidikan formal di TK PGRI Rejosari Bantur (1998-1999), SD Rejosari 1 Bantur (1999-2000), SD Oesapa Kecil 1 Kota Kupang (2000-2005), SMP Negeri 2 Kota Kupang (2005-2007), dan SMA Negeri 1 Kota Kupang (2007-2009). Penulis kemudian melanjutkan pendidikan ke jenjang S1 melalui Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) dan diterima di Prodi Statistika Universitas Brawijaya Malang (2009-2013). Setelah menyelesaikan studi S1 di tahun 2013, pada tahun yang sama penulis meneruskan studi ke jenjang S2 di Program Pascasarjana Statistika FMIPA Institut Teknologi Sepuluh Nopember Surabaya dengan biaya dari Beasiswa Pendidikan Pascasarjana Dalam Negeri (BPP DN) Dikti.

Saran, kritik, dan pertanyaan seputar tesis ini dapat disampaikan ke alamat email [evellinlusiana@gmail.com](mailto:evellinlusiana@gmail.com)